4.7 Hypothesis Testing

Mark Schemes

Course	DP IB Maths
Section	4. Statistics & Probability
Topic	4.7 Hypothesis Testing
Difficulty	Medium

Exam Papers Practice

To be used by all students preparing for DP IB Maths AI SL
Students of other boards may also find this useful

Question 1

a) H_0: The favourite music genre is independent of gender.

H_1: The favourite music genre is not independent of gender.

b) Degrees of freedom formula

$df = (c-1)(r-1)$ (not in formula booklet)

where c is the no. of columns and

r is the no. of rows.

$c = 4 \quad r = 2$

$df = (4-1)(2-1)$

$df = 3$

c) Create a 2x4 matrix on your GDC and perform a χ^2 2-way test.

$$\begin{pmatrix} 58 & 63 & 17 & 44 \\ 23 & 96 & 12 & 87 \end{pmatrix}$$

$\chi^2 = 33.9844$

$\chi^2 = 34.0$ (3sf)

Exam Papers Practice

d) $\chi^2 = 34.0$ $\chi^2_{cv} = 7.815$

$\chi^2 > \chi^2_{cv}$

Reject H_0, there is sufficient evidence to suggest that favourite music genre is dependent on gender.

Question 2

a) $H_0: \mu_1 - \mu_2 = 0$

The mean length of the pine needles at low altitude, μ_1, is the same as the mean of the pine needles at high altitude, μ_2.

$H_1: \mu_1 - \mu_2 \neq 0$

The mean length of the pine needles at low altitude, μ_1, is not the same as the mean of the pine needles at high altitude, μ_2.

Exam Papers Practice

b) One-tailed test when

$H_0: \mu_1 - \mu_2 = 0$

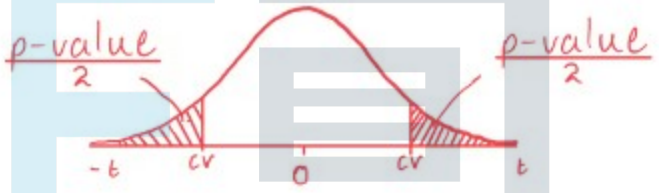
$H_1: \mu_1 - \mu_2 > 0$ or $\mu_1 - \mu_2 < 0$

Two-tailed test when

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 \neq 0$

$H_0: \mu_1 - \mu_2 = 0$ $H_1: \mu_1 - \mu_2 \neq 0$



∴ This is a two-tailed test.

c) Input the data into your GDC and perform a 2-Sample t-test.

List 1: Low altitude

List 2: High altitude

p-value = 0.8711...

p-value = 0.871 (3sf)

d) $p = 0.871$ $SL = 0.1$

$p > SL$

Do not reject H_0, there is insufficient evidence to suggest that the means are different.

Question 3

a) Expected value formula

$E(X) = \frac{\sum_{i=1}^{k} f_i x_i}{n}$ and $n = \sum_{i=1}^{k} f_i$ (in formula booklet)

$E(X) = \frac{16 + 12 + 14 + 20 + 15 + 19}{6}$

$E(X) = 16$ carpets per month

b) H_0: The data has a uniform distribution.

H_1: The data does not have a uniform distribution.

c) Degrees of freedom formula

$df = n - 1$ (not in formula booklet)

$df = 6 - 1$

$df = 5$

Exam Papers Practice

d) Input the data into your GDC and perform a χ^2 GOF test.

List 1: Observed list

16, 12, 14, 20, 15, 19

List 2: Expected list

16, 16, 16, 16, 16, 16

$df = 5$

p-value = 0.7192...

p-value = 0.719 (3sf)

e) $p = 0.719$ $SL = 0.05$

$p > SL$

Do not reject H_0, there is insufficient evidence to suggest that the number of carpets sold each month is not uniformly distributed.

Exam Papers Practice

Question 4

a) Expected value formula

$$E(X) = \frac{\sum_{i=1}^{k} f_i x_i}{n} \quad \text{and} \quad n = \sum_{i=1}^{k} f_i \quad \text{(in formula booklet)}$$

$$E(X) = \frac{473 + 405 + 512 + 467 + 503}{5}$$

$E(X) = 472$ applications

b)

H_0 : The data has a uniform distribution.

H_1 : The data does not have a uniform distribution.

c) Input the data into your GDC and perform a χ^2 GOF test.

List 1: Observed list

473, 405, 512, 467, 503

List 2: Expected list

472, 472, 472, 472, 472,

$df = 5 - 1 = 4$

p-value = 0.004718...

$\chi^2 = 14.9915...$

i) $\chi^2 = 15.0$ (3sf)

ii) p-value = 0.00472 (3sf)

Exam Papers Practice

d) $p = 0.00472$ $SL = 0.1$

$\chi^2 = 15.0$ $\chi^2_{cv} = 7.779$

$p < SL$ and $\chi^2 > \chi^2_{cv}$

$\therefore$ Reject H_0, the data does not fit a uniform distribution.

Question 5

a) H_0: Hair colour is independent of eye colour.

H_1: Hair colour is not independent of eye colour.

Exam Papers Practice

b) Degrees of freedom formula

$df = (c - 1)(r - 1)$ (not in formula booklet)

where c is the no. of columns and

r is the no. of rows.

$c = 3$ $r = 4$

$df = (3 - 1)(4 - 1)$

$df = 6$

c) $p = 0.0726$ $SL = 0.1$

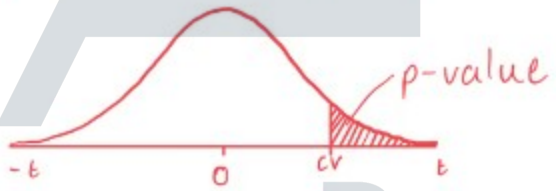
$p < SL$

$\therefore$ Reject H_0, there is sufficient evidence to suggest that hair colour is not independent of eye colour.

Question 6

a) "... women from Japan are *taller on average than woman from India"

$\therefore H_0: \mu_1 - \mu_2 = 0$ $H_1: \mu_1 - \mu_2 > 0$



where μ_1 = average height of Japanese women and μ_2 = average height of Indian women.

$\therefore$ This is a one-tailed t-test.

*This would be a two-tailed t-test if the claim was that the average heights of the women were different.

b) $H_0: \mu_1 - \mu_2 = 0$

The mean height of women from Japan, μ_1, is the same as the mean height of women from India, μ_2.

$H_1: \mu_1 - \mu_2 > 0$

The mean height of women from Japan, μ_1, greater than the mean height of women from India, μ_2.

c) Input the data into your GDC and perform a 2-Sample t-test.

List 1: Japan

List 2: India

p-value = 0.03033135...

p-value = 0.0303 (3sf)

Exam Papers Practice

d) $p = 0.0303$ $SL = 0.05$

$p < SL$

$\therefore$ Reject H_0, there is sufficient evidence to suggest that the mean height of Japanese women is greater than the mean height of Indian women.

Question 7

a) $H_0: \mu_1 - \mu_2 = 0$

The mean weight of a newborn at 38 weeks, μ_1, is the same as the mean weight of a newborn at 40 weeks, μ_2.

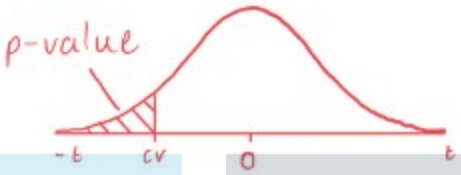
$H_1: \mu_1 - \mu_2 < 0$

The mean weight of a newborn at 38 weeks, μ_1, is less than the mean weight of a newborn at 40 weeks, μ_2.

Exam Papers Practice

b) "... average weight of a newborn baby born at 38 weeks is expected to be <u>less</u> than the average weight of a newborn baby born at full term (40 weeks)"

$\therefore H_0: \mu_1 - \mu_2 = 0 \quad H_1: \mu_1 - \mu_2 < 0$



where μ_1 = average weight of a 38 week newborn and μ_2 = average weight of a 40 week newborn.

∴ This is a one-tailed t-test.

c) Input the data into your GDC and perform a 2-Sample t-test.

List 1: 38 weeks

List 2: 40 weeks

p-value = 0.001245...

p-value = 0.00125 (3sf)

Exam Papers Practice

d) $p = 0.00125$ $SL = 0.1$

$p < SL$

$\therefore$ Reject Ho, there is sufficient evidence to suggest that the mean weight of a newborn baby born at 38 weeks is less than the mean weight of a newborn baby born at 40 weeks.

Question 8

	Red	Yellow	Blue	Green	Pink	Total
Children	17.5	11.8	18.3	13.9	10.5	72
Teenagers	20.6	14.0	21.6	16.4	12.4	85
Adults	29.9	20.2	31.2	23.7	18.0	123
Total	68	46	71	54	41	280

Exam Papers Practice

a) Expected values for contingency tables.

$$E(x) = \frac{\text{row total}}{\text{total}} \times \frac{\text{column total}}{\text{total}} \times \text{total}$$

For example:

Expected value of children who prefer red.

$$E(x) = \frac{72}{280} \times \frac{68}{280} \times 280$$

$$E(x) = 17.5$$

b)

H_0 : Favourite colour is independent of age.

H_1 : Favourite colour is not independent of age.

c) Degrees of freedom formula

$df = (c-1)(r-1)$ (not in formula booklet)

where c is the no. of columns and r is the no. of rows.

$c = 5 \quad r = 3$

$df = (5-1)(3-1)$

$df = 8$

Exam Papers Practice

d) Create a 3×5 matrix on your GDC and perform a χ^2 2-way test.

$$\begin{pmatrix} 20 & 11 & 18 & 8 & 15 \\ 22 & 14 & 23 & 20 & 6 \\ 26 & 21 & 30 & 26 & 20 \end{pmatrix}$$

$\chi^2 = 10.138$

$\chi^2 = 10.1$ (3sf)

c) $\chi^2 = 10.136$ $\chi^2_{cv} = 15.507$

$\chi^2 < \chi^2_{cv}$

$\therefore$ Do not reject H_0, there is insufficient evidence to suggest that the choice of favourite colour is dependent of age.

Question 9

Height (cm)	Probability	Expected frequency
$h < 3$	0.126549	44.3
$3 \leq h < 4$	0.485902	170.1
$4 \leq h < 5$	0.34431	120.5
$5 \leq h < 6$	0.042401	14.8

a) Calculate the probabilities using the normal CDF function on your GDC.

Calculate the expected frequencies by multiplying the probabilities by 350.

b) H_0: The heights of the giraffes fit a normal distribution.

H_1: The heights of the giraffes do not fit a normal distribution.

Exam Papers Practice

c) Degrees of freedom formula

$df = n - 1$ (not in formula booklet)

$df = 4 - 1$

$df = 3$

d) Input the data into your GDC and perform a X^2 GOF test.

List 1: Observed list

50, 160, 119, 21

List 2: Expected list

44.3, 170.1, 120.5, 14.8

$df = 3$

p-value = 0.26701...

p-value = 0.267 (3sf)

e) $p = 0.267$ $SL = 0.1$

$p > SL$

∴ Do not reject Ho, there is not sufficient evidence to suggest that the heights of these giraffes does not fit a normal distribution.

Exam Papers Practice