**IB Math Studies – Two Variable Statistics – Review Questions**

**1.** The following table of observed results gives the number of candidates taking a Mathematics examination classified by gender and grade obtained.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **Grade** |  |  |
|  |  | 5, 6 or 7 | 3 or 4 | 1 or 2 | Total |
|  | Males | 5000 | 3400 | 600 | 9000 |
| **Gender** | Females | 6000 | 4000 | 1000 | 11000 |
|  | Total | 11000 | 7400 | 1600 | 20000 |

The question posed is whether gender and grade obtained are independent.

(a) **Show clearly** that the expected number of males achieving a grade of 5, 6 or 7 is 4950.

(2)

(b) A  test is set up.

(i) State the Null hypothesis.

(1)

(ii) State the number of degrees of freedom.

(1)

(iii) The calculated  value at the 5% test level is 39.957.   
Write down the critical value of  at the 5% level of significance.

(1)

(iv) What can you say about gender and grade obtained?

(1)

(Total 6 marks)

**2.** A researcher consulted 500 men and women to see if the colour of the car they drove was independent of gender. The colours were red, green, blue, black and silver. A  test was conducted at the 5% significance level and the value found to be 8.73.

(a) Write down the null hypothesis.

(b) Find the number of degrees of freedom for this test.

(c) Write down the critical value for this test.

(d) Is car colour independent of gender? Give a clear reason for your answer

(Total 6 marks)

**3.** In a competition the number of males and females taking part in different swimming races is given in the table of observed values below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Backstroke (100 m) | Freestyle (100 m) | Butterfly (100 m) | Breaststroke (100 m) | Relay (4 × 100 m) |
| Male | 30 | 90 | 31 | 29 | 20 |
| Female | 28 | 63 | 20 | 37 | 12 |

The Swimming Committee decides to perform a *χ*2 test at the 5 significance level in order to test if the number of entries for the various strokes is related to gender.

(a) State the null hypothesis.

(1)

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

(1)

(c) Write down the critical value of *χ*2.

(1)

The expected values are given in the table below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Backstroke (100 m) | Freestyle (100 m) | Butterfly (100 m) | Breaststroke (100 m) | Relay (4 × 100 m) |
| Male | 32 | *a* | 28 | 37 | 18 |
| Female | 26 | 68 | 23 | *b* | 14 |

(d) Calculate the values of *a* and*b*.

(2)

(e) Calculate the *χ*2 value.

(3)

(f) State whether or not you accept your null hypothesis and give a reason for your answer.

(2)

(Total 10 marks)

**4.** The *Type Fast* secretarial training agency has a new computer software spreadsheet package. The agency investigates the number of hours it takes people of varying ages to reach a level of proficiency using this package. Fifteen individuals are tested and the results are summarised in the table below.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | 32 | 40 | 21 | 45 | 24 | 19 | 17 | 21 | 27 | 54 | 33 | 37 | 23 | 45 | 18 |
| (*x*) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Time |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (in hours) | 10 | 12 | 8 | 15 | 7 | 8 | 6 | 9 | 11 | 16 | *t* | 13 | 9 | 17 | 5 |
| (*y*) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

(a) (i) Given that *Sy* = 3.5 and *Sxy* = 36.7, calculate the product-moment correlation coefficient *r* for this data.

(4)

(ii) What does the value of the correlation coefficient suggest about the relationship between the two variables?

(1)

(b) Given that the mean time taken was 10.6 hours, write the equation of the regression line for *y* on *x* in the form *y* = *ax* + *b*.

(3)

(c) Use your equation for the regression line to predict

(i) the time that it would take a 33 year old person to reach proficiency, giving your answer correct to the nearest hour;

(2)

(ii) the age of a person who would take 8 hours to reach proficiency, giving your answer correct to the nearest year.

(2)

(Total 12 marks)

**5.** Ten students were asked for their average grade at the end of their last year of high school and their average grade at the end of their last year at university. The results were put into a table as follows:

|  |  |  |
| --- | --- | --- |
| Student | High School grade, *x* | University grade, *y* |
| 1  2  3  4  5  6  7  8  9  10 | 90  75  80  70  95  85  90  70  95  85 | 3.2  2.6  3.0  1.6  3.8  3.1  3.8  2.8  3.0  3.5 |
| Total | 835 | 30.4 |

(a) Given that *sx* = 8.96, *sy* = 0.610 and *sxy* = 4.16, find the correlation coefficient *r*, giving your answer to two decimal places.

(2)

(b) Describe the correlation between the high school grades and the university grades.

(2)

(c) Find the equation of the regression line for *y* on *x* in the form *y = ax + b*.

(2)

(Total 6 marks)

**6.** A study was carried out to investigate possible links between the weights of baby rabbits and their mothers. A sample of 20 pairs of mother rabbits (*x*) and baby rabbits (*y*) was chosen at random and their weights noted. This information was plotted on a scatter diagram and various statistical calculations were made. These appear below.



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| mean of *x* | mean of *y* | *sx* | *sy* | *sxy* | sum of *x* | sum of *y* |
| 3.78 | 3.46 | 0.850 | 0.689 | 0.442 | 75.6 | 69.2 |

(a) Show that the product-moment correlation coefficient *r* for this data is 0.755.

(2)

(b) (i) Write the equation of the regression line for *y* on *x* in the form *y* = *ax* + *b*.

(3)

(ii) Use your equation for the regression line to estimate the weight of a rabbit given that its mother weighs 3.71 kg.

(2)

(Total 7 marks)

**Mark Scheme**

**1.** (a) Males =  (M1)(A1)  
= 4950 (AG) 2

(b) (i) That gender and grade obtained are independent. (A1)  
(There is no connection between gender and grade obtained.)

(ii) (3 – 1)(2 – 1) = 2 (A1)

(iii) **2 =5.991 (A1)

(iv) Calculated**2 = 39.957  
Therefore, reject the Null hypothesis. Gender and grade obtained (R1) 4  
are dependent (or there is a connection between gender and grade).

[6]

**2.** (a) Colour of car and gender are independent (A1) (C1)

(b) (2 – 1) (5 – 1) (M1)  
= 4 (A1)

**OR**4 (A2) (C2)

(c) **2 = 9.488 (A1) (C1)

(d) Yes. Test statistic is smaller than the critical value. (A1)(R1) (C2)

[6]

**3.** (a) H0 : number of entries is independent of gender. (A1) 1

(b) 4 (A1) 1

(c) 9.488 (A1) 1

(d) *a* = 85, *b* = 29 (A1)(A1) 2

(e)  (M1)(A1)

= 6.10 (using given values) (A1)

**OR**

5.80 (from calculator) (G3) 3

(f) Do not reject the null hypothesis as the 2 value is less than the critical value.  
So, gender and stroke are independent. (A1)(R1)2

(Also allow “accept”).

[10]

**4.** (a) (i) *Sx* = 11.2 (A1)  
*r* =  (M2)  
 = 0.936 (3 s.f.) (A1)

**OR**

*Sx* = 11.6 (A1)  
*r* =  (M2)  
 = 0.904 (3 s.f.) (A1)

(ii) The correlation coefficient suggests a strong positive correlation  
between the two variables. (R1) 5

(b) *y* –   
*y* – 10.6 = (*x* – 30.4) (M1)  
*y* = 0.293*x* + 1.69 (or *y* = 0.293*x* + 1.71) *Allow ft from (a) (i))* (A2) 3

(c) (i) *y* = 0.293 × 33 + 1.69 (M1)  
= 11.359  
= 11 hours (A1)

(ii) 8 = 0.293*x* + 1.69 (M1)  
*x* = 21.54  
 = 22 years (A1) 4

[12]

**5.** (a) *r* =   
=  (M1)  
= 0.76 (A1) 2

(b) There is a fairly strong positive correlation between high school  
grades and university grades. (A1) (A1) 2

**Note:** Award (A1) for strong (or fairly strong) or high, (A1) for positive.

(c) *y* –   
*y* – 3.04 = (*x* – 83.5) (M1)  
*y* = 0.052*x* – 1.29 (3 s.f.) (A1) 2

**Note:** Award (C2) for correct answer (from calculator).

[6]

**6.** (a) *r* =  = 0.755 (M1)(M1)(AG) 2

(b) (i) *y* – ; *y* – 3.46 = (*x* – 3.78) (M1)  
*y* = 0.612*x* + 1.15 (A1)(A1)

(ii) weight of rabbit = 0.612 × 3.71 + 1.15 = 3.42 kg (M1)(A1) 5

[7]