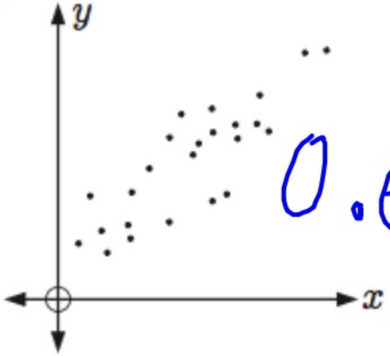
IB Math Studies 2 BELL WORK

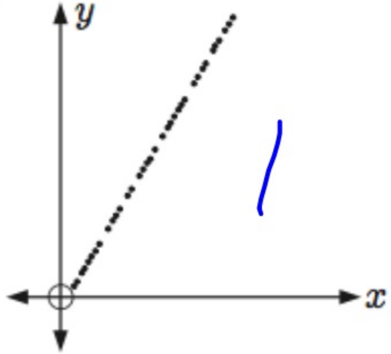
Match each scatter diagram with the correct value of r.

a



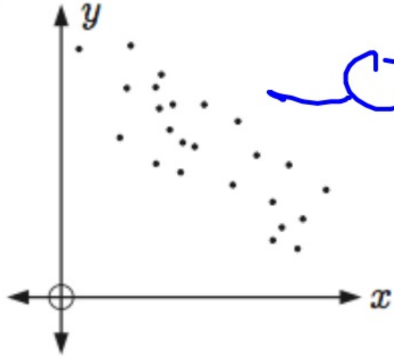
0.6

b



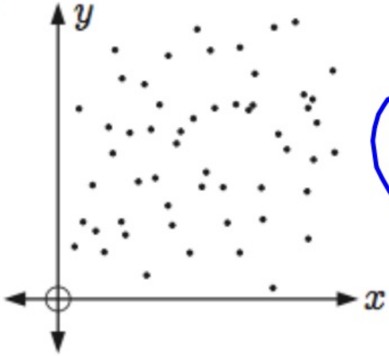
1

c



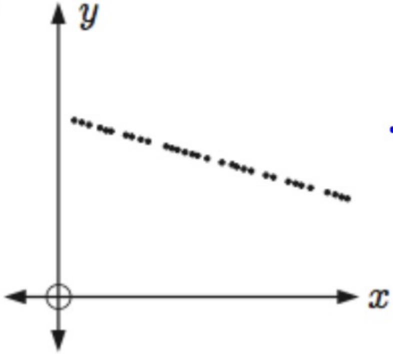
-0.7

d



0

e



-1

A $r = 1$ **B** $r = 0.6$ **C** $r = 0$ **D** $r = -0.7$ **E** $r = -1$

E THE χ^2 TEST OF INDEPENDENCE

continued...

* How to find the chi-squared statistic on the calculator

* How to use this statistic to make conclusions about your data
  -The Formal Test for Independence

* Other considerations when looking at the data results

FORMAL TEST FOR INDEPENDENCE

We have seen that a small value of χ^2 indicates that two variables are independent, while a large value of χ^2 indicates that the variables are not independent.

We will now consider a more formal test which determines *how large* χ^2 must be for us to conclude the variables are not independent.

This is known as the **critical value** of χ^2.

The critical value of χ^2 depends on:

- the **size** of the contingency table, measured by **degrees of freedom**
- the **significance level** used.

DEGREES OF FREEDOM

In a contingency table, the number of **degrees of freedom (df)** is the number of values which are free to vary.

Consider a 2x2 contigency table with the sum values given, as shown. In this table, the first number is free to be any number less than the sum value.

However, one that number is set, the other numbers must take specific values in order to give the sums.

So the degrees of freedom is 1, which is (2-1) x (2-1)

	A_1	A_2	*sum*
B_1	X		12
B_2			8
sum	15	5	20

For a contingency table which has r rows and c columns, $\mathbf{df} = (\boldsymbol{r} - \mathbf{1})(\boldsymbol{c} - \mathbf{1})$.

SIGNIFICANCE LEVEL

As the χ^2 value gets larger, it becomes more likely that the variables are not independent. The **significance level** indicates the minimum acceptable probability that the variables are independent.

95%

We usually use either 10%, 5%, or 1% for the significance level.

90% certain 99%

For a given significance level and degrees of freedom, the table alongside gives the **critical value** of χ^2, above which we conclude the variables are not independent.

For example, at a 5% significance level with $\text{df} = 1$, the critical value is 3.84. This means that at a 5% significance level, the departure between the observed and expected values is too great if $\chi^2_{calc} > 3.84$.

Degrees of	*Significance level*		
freedom (df)	10%	5%	1%
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
	14.68	16.92	21.67
	15.99	18.31	23.21

Likewise, at a 1% significance level with $\text{df} = 7$, the departure between the observed and expected values is too great if $\chi^2_{calc} > 18.48$.

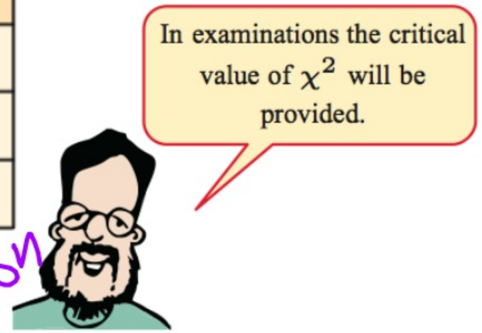
How to use the calculator to find *p*-value and Chi-Squared Statistic

Hockey player Julie wondered whether the position you played affected your likelihood of being injured. She asked a random sample of hockey players what position they played, and what injuries they had sustained in the last year.

Injury type	*Position*: *Forward*	*Midfielder*	*Defender*	*Goalkeeper*
No injury	23	18	24	7
Mild injury	14	34	23	11
Serious injury	10	16	13	7

Test, at a 10% significance level, whether the variables *position* and *injury type* are independent.

		Position			
		Forward	*Midfielder*	*Defender*	*Goalkeeper*
Injury type	*No injury*	23	18	24	7
	Mild injury	14	34	23	11
	Serious injury	10	16	13	7



Degrees of freedom (df)	*Significance level*		
	10%	5%	1%
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21

$\chi^2_{CALC} = 7.942$

p-value = 0.242

df = 6

$\chi^2_{CRIT} = 10.64$

low χ^2 so position and injury are independent

THE p-VALUE

When finding χ^2 on your calculator, a ***p*-value** is also provided.

For a given contingency table, the ***p*-value** is the probability of obtaining observed values as far or further from the expected values, assuming the variables are independent.

It is not always essential to use the p-value when testing for independence, as we can perform the test by simply comparing χ^2_{calc} with the critical value. However, the p-value does give a more meaningful measure of how likely it is that the variables are independent.

If the p-value is smaller than the significance level, then it is sufficiently unlikely that we would have obtained the observed results if the variables had been independent. We therefore conclude that the variables are not independent.

THE FORMAL TEST FOR INDEPENDENCE

Step 1: State H_0 called the **null hypothesis**. This is a statement that the two variables being considered are independent.

State H_1 called the **alternative hypothesis**. This is a statement that the two variables being considered are not independent.

Step 2: State the **rejection inequality** $\chi^2_{calc} > k$ where k is the **critical value** of χ^2.

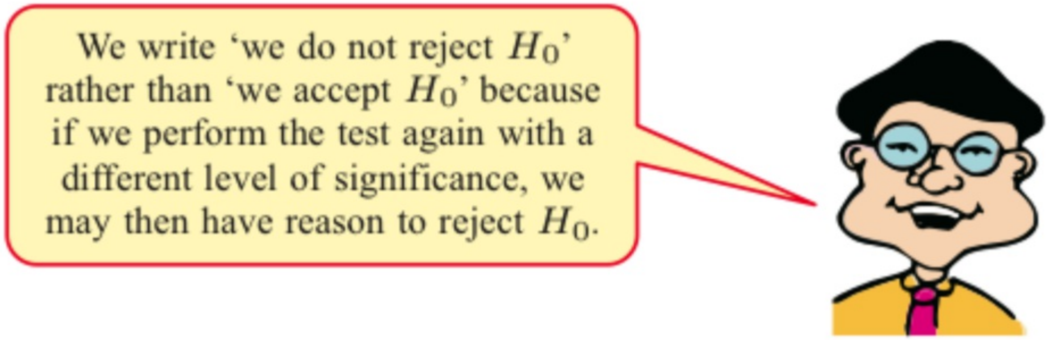
Step 3: Construct the expected frequency table.

Step 4: Use technology to find χ^2_{calc}.

Step 5: We either reject H_0 or do not reject H_0, depending on the result of the rejection inequality.

Step 6: We could also use a ***p*-value** to help us with our decision making.

For example, at a 5% significance level: If $p < 0.05$, we reject H_0.
If $p > 0.05$, we do not reject H_0.



A survey was given to randomly chosen high school students from years 9 to 12 on possible changes to the school's canteen.

The contingency table shows the results.

At a 5% significance level, test whether the student's *canteen preference* depends on the year group.

	Year group			
	9	10	11	12
change	7	9	13	14
no change	14	12	9	7

χ^2_{CRIT} : 7.81

χ^2_{CALC} : 5.812

χ^2-Test

df: (2-1)(4-1)

1 · 3

p : 0.121

df: 3

Assignment:

Exercise 11 E.2 #2, 3, 4