IB Math Studies 2

Grab a copy of the notes from the chair up front.

(Being passed out)

E THE χ^2 TEST OF INDEPENDENCE

The **chi-squared** or χ^2 **test** is used to determine whether two variables from the same sample are independent.

independent: the occurrence of one factor ***does not affect*** the occurrence of the other

dependent: the occurrence of one factor ***affects*** the occurrence of the other

The Chi Squared Test is applied when you have two categorical variables from a single population. It is used to determine whether there is a significant association between the two variables.

For example, in an election survey, voters might be classified by gender (male or female) and voting preference (Democrat, Republican, or Independent). We could use a chi-square test for independence to determine whether gender is related to voting preference.

http://stattrek.com/chi-square-test/independence.aspx?Tutorial=AP

When to Use Chi-Square Test for Independence

The Chi squared test is appropriate when the following conditions are met:

- The sampling method is simple random sampling.
- The variables under study are each categorical.
- If sample data are displayed in a contingency table, the *expected* frequency count for each cell of the table is at least 5. (less is not acceptable)

http://stattrek.com/chi-square-test/independence.aspx?Tutorial=AP

This approach consists of four steps:

(1) state the hypotheses,
(2) formulate an analysis plan,
(3) analyze sample data, and
(4) interpret results.

http://stattrek.com/chi-square-test/independence.aspx?Tutorial=AP

State the Hypotheses

Suppose that Variable A has *r* levels, and Variable B has *c* levels. The null hypothesis states that knowing the level of Variable A does not help you predict the level of Variable B. That is, the variables are independent. H_0 null: not dependent

H_A alternate: dependent

The alternative hypothesis is that knowing the level of Variable A *can* help you predict the level of Variable B.

Note: Support for the alternative hypothesis suggests that the variables are related; but the relationship is not necessarily causal.

http://stattrek.com/chi-square-test/independence.aspx?Tutorial=AP

Formulate an Analysis Plan

The analysis plan describes how to use sample data to accept or reject the null hypothesis. The plan should specify the following elements.

- Significance level. Often, researchers choose significance levels equal to 0.01, 0.05, or 0.10; but any value between 0 and 1 can be used.

- Test method. Use the chi-square test for independence to determine whether there is a significant relationship between two categorical variables.

http://stattrek.com/chi-square-test/independence.aspx?Tutorial=AP

Analyze Sample Data

Using sample data, find the degrees of freedom, expected frequencies, test statistic, and the P-value associated with the test statistic.

Degrees of freedom. The degrees of freedom (DF) is equal to: (rows - 1)(columns - 1)

$$DF = (r - 1) * (c - 1)$$

where r is the number of levels for one catagorical variable, and c is the number of levels for the other categorical variable.

http://stattrek.com/chi-square-test/independence.aspx?Tutorial=AP

Expected frequencies.

The expected frequency counts are computed separately for each level of one categorical variable at each level of the other categorical variable. Compute r * c expected frequencies, according to the following formula.

$$f_e = \frac{(row\ total)(column\ total)}{table\ total}$$

http://stattrek.com/chi-square-test/independence.aspx?Tutorial=AP

Test statistic. χ^2_{CALC}

The test statistic is a chi-square random variable (X^2) defined by the following equation.

$$\chi^2 = \sum \left[\frac{(f_o - f_e)^2}{f_e} \right]$$

where

f_o is the observed frequency, and

f_e is the expected frequency.

http://stattrek.com/chi-square-test/independence.aspx?Tutorial=AP

P-value.

The P-value is the probability of observing a sample statistic as extreme as the test statistic.

You will find this value using your graphing calculator.

http://stattrek.com/chi-square-test/independence.aspx?Tutorial=AP

Interpret Results

If the sample findings are unlikely, given the null hypothesis, the researcher rejects the null hypothesis. Typically, this involves comparing the P-value to the significance level, and rejecting the null hypothesis when the P-value is less than the significance level.

http://stattrek.com/chi-square-test/independence.aspx?Tutorial=AP

If the variables are independent, the observed and expected values will be very similar. This means that the values of $(f_o - f_e)$ will be small, and hence χ^2_{calc} will be small.

If the variables are not independent, the observed values will differ significantly from the expected values. The values of $(f_o - f_e)$ will be large, and hence χ^2_{calc} will be large.

Now let's try it:

This table shows the results of a sample of 400 randomly selected adults classified according to *gender* and *regular exercise*.

	Regular exercise	*No regular exercise*	*sum*
Male	110	106	216
Female	98	86	184
sum	208	192	400

Remember, a contingency table is a FREQUENCY table

For the expected values, we look at the totals in the contingency table and determine what we would expect to be in each column if the variables were independent.

	Regular exercise	*No regular exercise*	*sum*
Male	$\frac{(216)(208)}{400}$	$\frac{(216)(192)}{400}$	216
Female	$\frac{(184)(208)}{400}$	$\frac{(184)(192)}{400}$	184
sum	208	192	400

	Regular exercise	*No regular exercise*	*sum*
Male	112.32	103.68	216
Female	95.68	88.32	184
sum	208	192	400

	Regular exercise	*No regular exercise*	*sum*
Male			216
Female			184
sum	208	192	400

For example, if *gender* and *regular exercise* were independent, then

$$\begin{aligned} \mathrm{P}(\textit{male} \cap \textit{regular exercise}) &= \mathrm{P}(\textit{male}) \times \mathrm{P}(\textit{regular exercise}) \\ &= \frac{216}{400} \times \frac{208}{400} \end{aligned}$$

So, in a sample of 400 adults, we would expect

$$400 \times \left(\frac{216}{400} \times \frac{208}{400}\right) = \frac{216 \times 208}{400} = 112.32$$ to be male and exercise regularly.

	Regular exercise	*No regular exercise*	*sum*
Male			216
Female			184
sum	208	192	400

Complete the ***expected frequency*** *table.*

The χ^2 test examines the difference between the **observed** values we obtained from our sample, and the **expected** values we have calculated.

$$\chi^2_{calc} = \sum \frac{(f_o - f_e)^2}{f_e}$$

where f_o is an observed frequency
and f_e is an expected frequency.

Set up a table to calculate your chi-squared statisitc.

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
110	112.32	2.32	5.3824	0.047920279
106	103.68	2.32	5.3824	0.051913582
98	95.68	2.32	5.3824	0.056254806
86	88.32	2.32	5.3824	0.060942029
			Total	0.217134

0.217

Test Your Understanding

Problem

A public opinion poll surveyed a simple random sample of 1000 voters. Respondents were classified by gender (male or female) and by voting preference (Republican, Democrat, or Independent). Results are shown in the contingency table below.

	Voting Preferences			Row total
	Republican	Democrat	Independent	
Male	200	150	50	400
Female	250	300	50	600
Column total	450	450	100	1000

Is there a gender gap? Do the men's voting preferences differ significantly from the women's preferences? Use a 0.05 level of significance.

Assignment:

Exercise 11 E.1 #2

Consider the contingency table:

	Pass Maths test	*Fail Maths test*	*sum*
Male	24	26	50
Female	36	14	50
sum	60	40	100

a Construct an expected frequency table.

b Interpret the value in the top left corner of the expected frequency table.

c Calculate χ^2_{calc} by copying and completing this table:

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
24				
26				
36				
14				
			Total	