

Statistical Applications

4: χ^2 Test of Independence

Chapter 5: p233 - 241

χ^2 Test

- The χ^2 (chi squared) test is used to determine if two factors from the same sample are independent.
- Independence means that the occurrence of one of them does not affect the occurrence of the other.
- Examples of two factors:
 - Income & voting intentions
 - Gender & money earning potential
 - Social networking & GPA
 - Gender & handedness

1: Null Hypothesis

- The formal test for independence starts with a statement of the null hypothesis (H_0) which states that the data is independent. ^{must use this word}
- At the same time, an alternative hypothesis (H_1) is made, which states that the data is not independent.

2: Observed Values

- A contingency table is constructed. This shows the observed values collected. Results from survey/website

Table of Observed Values

	Grade 9	Grade 10	Grade 11	TOTALS
Left handed	6	9	13	28
Right handed	73	92	67	232
TOTALS	79	101	80	260

- The table shows 260 randomly selected children; sorted by grade level & handedness.

Probability for 2 Events

- The probability of two independent events happening together is equal to the product of the probabilities of each event.

Eg. Tossing a coin & throwing a dice at the same time (= independent events)



Probability of throwing a 6 = $\frac{1}{6}$

Probability of getting "heads" = $\frac{1}{2}$

Probability of getting both a 6 and "heads" at the same time

$$\frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$$

If the coin & dice are thrown together 36 times, the expected frequency of "heads + 6" = probability x number $\therefore \frac{1}{12} \times 36 = 3$ times

3: Expected Values

- If grade level is independent of handedness, the observed probability = expected probability.

expected probability of G9 student being LH
 $= \frac{79}{260} \times \frac{28}{260} = 0.0327$

	Grade 9	Grade 10	Grade 11	TOTALS
Left handed	6	9	13	28
Right handed	73	92	67	232
TOTALS	79	101	80	260

- expected probability x total number of children = expected value of LH grade 9 students:

$$0.0327 \times 260 = 8.51$$

Keep as a decimal

3: Expected Values

- Calculate all the expected values: ** expect this Q in an IB exam ∴ learn the method*

$$G9 \times RH: \frac{79}{260} \times \frac{232}{260} \times 260 = 70.5$$

$$G10 \times LH: \frac{101}{260} \times \frac{28}{260} \times 260 = 10.9$$

$$G10 \times RH: \frac{101}{260} \times \frac{232}{260} \times 260 = 90.1$$

$$G11 \times LH: \frac{80}{260} \times \frac{28}{260} \times 260 = 8.62$$

$$G11 \times RH: \frac{80}{260} \times \frac{232}{260} \times 260 = 71.4$$

Table of Expected Values

	Grade 9	Grade 10	Grade 11	TOTALS
Left handed	8.51	10.9	8.62	
Right handed	70.5	90.1	71.4	
TOTALS				260

4: Degrees of Freedom

- The χ^2 distribution is dependent on the number of degrees of freedom (df).

$$\text{degrees of freedom} = (\text{rows} - 1)(\text{columns} - 1)$$

- In our example; $df = (2 - 1) \times (3 - 1) = 2 df$

Using the Calculator

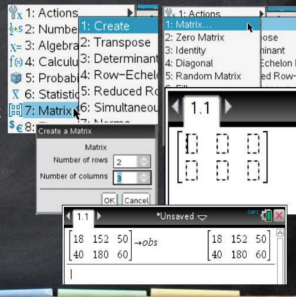
Entering observed values

- The eye colour and gender of 500 students are noted and the results are indicated in the table at right. Calculate the expected values and the χ^2 value for these results.

	Blue	Brown	Green
Male	18	152	50
Female	40	180	60

Open a calculate scratchpad:

- Press MENU → 7:Matrix & Vector → 1:Create → 1:Matrix
- Enter the number of rows & columns, OK
- A matrix is created for you to enter the observed values into. Use TAB to move around
- Make sure your cursor is outside the matrix, Press CTRL VAR and type "obs" press ENTER



Using the Calculator

Finding χ^2 Value and Expected Value

Open a calculate scratchpad:

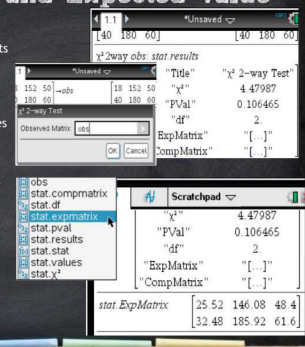
- Press MENU → 6:Statistics → 7:Stat Tests → 8: χ^2 2-way Test, ENTER
- Choose "obs" from menu, OK

- Information Given: The results are the value for χ^2 , the p value and the degrees of freedom.

To calculate Expected Values: You MUST calculate χ^2 first for this to work!

- Press VAR → 6:Statistics → stat.expmatrix, ENTER

- Information Given: The experimental values are in the same position as the observed values



5: Interpreting the χ^2 Test

- We compare either the χ^2 calculated value or the p-value to another number to see if the null hypothesis is accepted or rejected. ** next slides*

- If H_0 is accepted, the 2 variables are independent which means one variable has no effect on the other.

- If H_0 is rejected, the 2 variables are dependent and a change in one variable will affect the other.

p - Values

- p-values are a way to decide to accept or reject the null hypothesis. *0.95 (95%) confident that the difference between obs & exp is only due to chance.*

- If the p-value is more than the significance level, (usually 0.05) we accept the null hypothesis.

- If the p-value is less than the significance level, we reject the null hypothesis.

- Our handedness example: $p = 0.158$ which is more than the significance level of 0.05 so we accept the null hypothesis.

- We conclude that handedness is independent of grade level.

Critical Values

In tests you may be given a critical value to compare your χ^2 result to. **LEARN THIS**

▪ If the χ^2 value is less than the critical value, we accept the null hypothesis (ie the data is independent).

▪ If the χ^2 value is more than the critical value, we reject the null hypothesis (ie the data is dependent).

• Our handedness example: the critical value is 5.991 at 5% significance level.

• We calculated the χ^2 value = 3.69 which is less than 5.991, so we accept the null hypothesis. \therefore handedness is independent of grade level.

Example to try:

2. The table below shows the favourite colour of car for men and women.

	Black	White	Red	Blue	Total
Male	51	22	33	24	130
Female	45	36	22	27	130
Total	96	58	55	51	260

a) calculate the χ^2 value for this data.

b) complete the table to below to show the expected frequencies

	Black	White	Red	Blue	Total
Male	48	29	31.5	25.5	130
Female	48	29	23.5	25.5	130
Total	96	58	55	51	260

c) show that the expected frequency for females preferring white cars is 29.

$$\frac{130}{260} \times \frac{58}{260} \times 260 = 29$$

d) State the null hypothesis and the alternative hypothesis.

H_0 : Gender & favourite car colour are independent.

e) Should H_0 be accepted? Give a reason why.

$p = 0.105 > 0.05 \therefore$ we accept H_0

Practice

Ex 5H: p238 Q2 - 8 evens