

Maths Statistics Chi-squared Test: Subject Choices

# Chi-squared Test: Subject Choices



A group of students can choose two of three subjects: Physics, Chemistry and Biology. The number choosing each possible pairing and their gender is given in the table below.

	Subject choice					
	Physics/Chemistry	Chemistry/Biology	Physics/Biology	Total		
Male	210	28	111	349		
Female	76	17	93	186		
Total	286	45	204	535		

Test whether there is evidence for a gender bias in the subjects chosen at the 1% level of significance.

### Part A Gender balance: expected numbers

Some of the expected numbers have been filled in in the table below. Find the numbers a, b and c giving your answers to 4 s.f.

	Subject choice				
	Physics/Chemistry	s/Chemistry Chemistry/Biology Physics/Bio			
Male	186.6	b	133.1		
Female	a	15.64	c		

Find a, the expected number of female students doing Physics/Chemistry.

Find b, the expected number of male students doing Chemistry/Biology.

Find c the expected number of females doing Physics/Biology.

### Part B Gender bias: chi-squared value

Calculate the chi-squared statistic for the given data. Give your answer to  $4\ \mathrm{s.f.}$ 

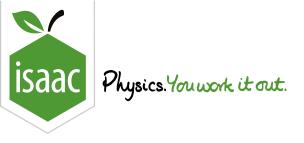
### Part C Gender bias: significance

Find the appropriate critical value of chi-squared at the 1% level of significance. Give your answer to 4 s.f.

## Part D Hypothesis test

What do you conclude about whether there is evilevel of significance?	dence for a gender bias in the subjects chosen at the $1\%$
$\mathrm{H}_0$ : Subject pairs and gender are $oxedsymbol{oxedge}$ .	
$\mathrm{H}_1:$ Subject pairs and gender are $oxedsymbol{oxedge}$ .	
The calculated value of chi-squared is	the critical value at a significance level of $1\%$ . Therefore at
this level we $igg[ H_0.$ There is $igg[$	evidence that subject pairs are not independent of gender.
Items:	
independent not independent do not reject	significant less than no significant reject
greater than	

Created for isaacphysics.org by Julia Riley



Maths Statistics Chi-squared Test: Handedness

# Chi-squared Test: Handedness



There is a suggestion that whether someone is right-handed or left-handed is correlated with their subject preferences. Groups of students in a college studying three different subjects: Mathematics, Geography and Law were asked whether they were right- or left-handed; the groups do not overlap. The numbers in each subject and their handedness is given in the table below.

	Subject choice				
	Mathematics Geography				
Right-handed	172	58	65		
Left-handed	13	4	8		

Test whether there is evidence for a relationship between the subject chosen and the handedness of the student at the 10%level of significance.

## Part A Expected frequencies

Some of the expected frequencies have been filled in in the table below. Find the frequencies  $a,\,b$  and c giving your answers to 4 s.f.

	Subject choice				
	Mathematics Geography				
Right-handed	a	57.16	67.30		
Left-handed	eft-handed 14.45		c		

Find the value of a.

F	-ir	٦d	the	va	lue	of	b.

						•	
_	ind	tr	$\sim$	1/0	$\Box$	$\sim$ t	•
	11 1( )			va		( )	•
•		٠.		v	$\sim$	$\sim$ .	$\sim$

## Part B Using a chi-squared test

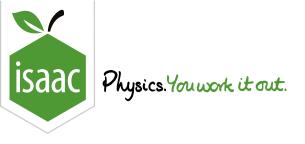
State the appropriate hypotheses for a chi-squared test on these data and any requirement this places on the expected frequencies.
$ m H_0:$ There is $ m igcirc$ for a relationship between the subject chosen and the handedness of a student.
$ m H_1:$ There is $ m igcirc$ for a relationship between the subject chosen and the handedness of a student.
The use of the chi-squared test requires that the frequencies in the contingency table must be at least. If this is not the case then the cells must be.
Items:
omitted     combined appropriately     no evidence     evidence     6     observed     4     7     3     expected
5

Part	C	The value of chi-squared
	do	the light of the expected values calculated in part A and the comments in part B, consider what you should to use a chi-squared test on this dataset. State the number of degrees of freedom after making any propriate alterations.
	Ca	lculate the value of the chi-squared statistic for this dataset. Give your answer to $4 { m s.f.}$
Part	D	The critical value of chi-squared.
	Fin	Id the appropriate critical value of chi-squared at the $10\%$ significance level. Give your answer to $4$ s.f.
Part	Wh	Hypothesis test  nat do you conclude about whether there is evidence for a relationship between the subject chosen and the
		ndedness of the student at the $10\%$ level of significance?
	at t	the critical value at a significance level of $10\%$ . Therefore this level we $H_0$ . There is evidence that subject choice and handedness are ated.
	Iter	ms:
		unrelated related reject significant no significant less than do not reject greater than

Created for isaacphysics.org by Julia Riley

Gameboard:

**STEM SMART Double Maths 46 - Chi-squared Tests** 



Maths Statistics

**Hypothesis Tests** 

Chi-squared Test: Biased Coins

# Chi-squared Test: Biased Coins



Three coins are tossed a number of times and the number of tails noted each time. It is suspected that one of the coins might be biased in the sense that heads and tails are not equally likely. Answer the following.

#### 32 tosses: chi-squared Part A

The coins are tossed 32 times and the following results achieved.

Number of tails	0	1	2	3
Frequency	6	14	6	6

It is assumed that the coins are unbiased so that it is equally likely that a head or tail will be obtained. Deduce the probability for each outcome and obtain the associated expected frequencies.

Calculate the chi-squared statistic for this dataset.

#### Part B 32 tosses: hypothesis test

Carry out a chi-squared test to determine at the 10% level whether the model fits the data for 32 tosses in part A.

Find the appropriate critical value of chi-squared. Give your answer to  $4\ \mathrm{s.f.}$ 

Using this critical value and the chi-squared statistic found in part A, what do you conclude about whether the model fits the data for $32$ tosses in part A?
$H_0$ : The coins are unbiased so that the frequencies of getting $0,1,2$ or $3$ tails in the ratio 1:
$H_1$ : The coins are biased so that the frequencies of getting $0,1,2$ or $3$ tails in the ratio 1:
The calculated value of chi-squared is $\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$
there is evidence that the ratios are different from 1: :1 and the coins are
biased.
Items:
are greater than no significant reject 4 significant less than 5 3 1 2
do not reject are not

### Part C 400 tosses, no bias: chi-squared

The sample considered above is rather small. To investigate further the three coins are tossed 400 times and the following results obtained.

Number of tails	0	1	2	3
Frequency	61	161	143	35

It is again assumed that they are unbiased so that it is equally likely that a head or tail will be obtained. Obtain the associated expected frequencies.

Calculate the chi-squared statistic for this dataset given the assumption above. Give your answer to  $4 \, \mathrm{s.f.}$ 

#### Part D 400 tosses, no bias: hypothesis test

Carry out a chi-squared test to determine at the 5% level whether the model fits the data for 400 tosses in part C.

Find the appropriate critical value of chi-squared at the 5% significance level. Give your answer to 4 s.f.

Using this critical value and the chi-squared statistic found in part C, what do you conclude about whether the model fits the data for $400$ tosses in part C?
$H_0$ : The coins are unbiased so that the frequencies of getting $0,1,2$ or $3$ tails in the ratio 1:
$H_1$ : The coins are biased so that the frequencies of getting $0,1,2$ or $3$ tails in the ratio 1:
The calculated value of chi-squared is $oxed{ ext{the critical value. Thus we}}$ the critical value. Thus we $oxed{ ext{H}_0}$ at the $5\%$ level;
there is evidence that the ratios are different from 1: :1 and the coins are
biased.
Items:  significant are 3 reject less than no significant do not reject 1 are not 2 5  greater than 4

### Part E 400 tosses, possible bias: chi-squared

The data obtained in Part C when the three coins are tossed 400 times is investigated further. The data is presented again below.

Number of tails	0	1	2	3
Frequency	61	161	143	35

It is now assumed that for one of the coins the probability of obtaining a tail is  $\frac{1}{3}$  rather than  $\frac{1}{2}$ ; the other two are unbiased. Obtain the associated expected frequencies in this case.

Calculate the chi-squared statistic for this dataset given the assumption above. Give your answer to  $4\ \mathrm{s.f.}$ 

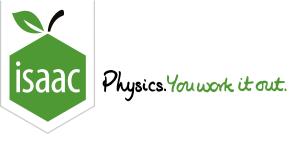
## Part F 400 tosses, possible bias: hypothesis test

Carry out a chi-squared test to determine at the $5\%$ level whether this model fits the data in part E.
$ m H_0$ : The coins are such that the frequencies of getting $0,1,2$ or $3$ tails $oxed{\hspace{0.5cm}}$ in the ratio $oxed{\hspace{0.5cm}}$ .
$ m H_1$ : The coins are such that the frequencies of getting $0,1,2$ or $3$ tails $oxed{\hspace{0.5cm}}$ in the ratio $oxed{\hspace{0.5cm}}$ .
The calculated value of chi-squared is $\begin{tabular}{ l l l l l l l l l l l l l l l l l l l$
Items:
2:3:2:1 significant do not reject 1:3:3:1 reject are less than no significant
2:4:5:1 2:5:4:1 more than are not

Created for isaacphysics.org by Julia Riley

Gameboard:

**STEM SMART Double Maths 46 - Chi-squared Tests** 



Maths

Statistics

Chi-squared Test: Active Galaxies

# Chi-squared Test: Active Galaxies



The number of galaxies with evidence for supermassive black holes in their centres are counted in 48 different independent areas of sky of equal area A. It is assumed that the number in each area follows a Poisson distribution. Using this assumption deduce the probability of finding that number and obtain the associated expected frequencies. Carry out a chisquared test to determine at the 10% level whether the data are consistent with the assumption that they come from a Poisson distribution.

Number of galaxies	0	1	2	3	4	5
Frequency	6	16	15	6	3	2

#### Part A The mean number of galaxies

From the observed data calculate the mean number of galaxies in area A. Give your answer to 4 s.f.

#### Part B Poisson distribution: chi-squared

It is assumed that the data can be modelled by a Poisson distribution with the mean calculated above. Using this assumption deduce the probability of finding each of the given numbers of galaxies and obtain the associated expected frequencies.

Calculate the value of the chi-squared statistic for this dataset. Give your answer to 3 s.f.

#### Part C Critical value: chi-squared

Find the critical value of chi-squared for a 10% level of significance. Give your answer to 3 s.f.

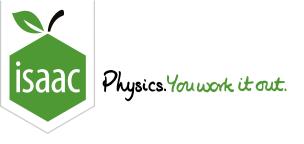
## Part D Hypothesis test

Carry out a chi-squared test to determine at the $10^\circ$ assumption that they come from a Poisson distribute	
$ m H_0:$ the galaxy distribution $oxedsymbol{oxedge}$ a Poisson of	distribution with mean as calculated.
$ m H_1$ : the galaxy distribution $ m igcap a$ Poisson of	distribution with mean as calculated.
The calculated value of chi-squared is	the critical value at a significance level of $10\%$ . Therefore
at this level we $oxedown$ $H_0$ . There is	evidence that a Poisson distribution fits the data.
Items:   reject greater than less than no significant	significant fits do not reject does not fit

Created for isaacphysics.org by Julia Riley

Gameboard:

**STEM SMART Double Maths 46 - Chi-squared Tests** 



Home Gameboard

iameboard Maths

Statistics

Chi-squared Test: Measurements

# Chi-squared Test: Measurements

Further AUniversity

Measurements of a particular quantity x can only take the values  $2, 4, 6, \dots 18$ .

A model for the expected frequencies is assumed which is symmetrical about 10.

1180 such measurements were made. The observed and some of the expected frequencies derived using the model (and given to the nearest integer) are in the table below.

Value	2	4	6	8	10	12	14	16	18
Observed frequency	45	106	122	198	206	205	158	77	63
Expected frequency	56	a	147	b	202	c	d	99	e

Carry out a chi-squared test to test the hypothesis that they follow the proposed distribution at (i) the 5% and (ii) 2.5% level.

Part /	Expected frequencies	
	educe the missing values in the table.	
	educe the value of $a$ .	
	educe the value of $d$ .	
	educe the value of $e$ .	
	nd the value of $b$ .	
	nd the value of $c$ .	
Part	The value of chi-squared	
	ssuming the integer expected frequencies deduced, calculate the chi-squared statistic for the given data. we your answer to $3\ \mathrm{s.f.}$	

# ${\bf Part \ C} \hspace{0.5cm} 5\% \ {\bf significance} \ {\bf level}$

Find the	critical	value fo	r chi-sq	uared us	sing a $5\%$	sigr 🌣	nificance	level.	Give y	your	answer	to 3	s.f	:
						~				,				

$H_1$ : The distribution	
this level we $oxedownder H_0$ . Th	vers is a vidence that the proposed model does not fit the data
	nere is evidence that the proposed model does not fit the data.
Items:	
greater than fits significan	nt no significant less than does not fit do not reject reject
g. cata. man	in and in the second of the se
•	whether the data follow the proposed distribution at the $2.5\%$ significance
ievei ?	
level? $H_0:$ the distribution	the proposed model.
	the proposed model.  the proposed model.
$ m H_0:$ the distribution	the proposed model.
$H_0:$ the distribution $H_1:$ the distribution $H_1:$ The calculated value of chi-square	the proposed model.
$H_0:$ the distribution $H_1:$ the distribution $H_1:$ The calculated value of chi-square	the proposed model.

Created for isaacphysics.org by Julia Riley