Chi-squared Test: Subject Choices

Further A University

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

In the table below some of the expected numbers have been filled in. Enter the missing numbers, giving your answers to 4 sf.

	Subject choice		
	Physics/Chemistry	**Chemistry/Biology**	**Physics/Biology**
Male	186.6		133.1
Female		15.64	

Part B Gender bias: chi-squared value

Calculate the chi-squared statistic for the given data. Give your answer to 4 sf.

Part C Gender bias: significance

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

Part D Hypothesis test

What do you conclude about whether there is evidence for a gender bias in the subjects chosen at the 1% level of significance?

H_0 : Subject pairs and gender are [ ].

H_1 : Subject pairs and gender are [ ].

The calculated value of chi-squared is [ ] the critical value at a significance level of 1%. Therefore at this level we [ ] H_0. There is [ ] evidence that subject pairs are not independent of gender.

Items:

greater than **reject** **no significant** **significant** **not independent** **independent** **do not reject** **less than**

Created for isaacphysics.org by Julia Riley

Chi-squared Test: Handedness

Further A University

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**	**Law**
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

In the table below some of the expected frequencies have been filled in. Enter the missing frequencies, giving your answers to 4 sf.

	Subject choice		
	Mathematics	**Geography**	**Law**
Right-handed		57.16	67.30
Left-handed	14.45		

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.

H_0 : There is [ ] for a relationship between the subject chosen and the handedness of a student.

H_1 : There is [ ] 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:

7 | 4 | 6 | **combined appropriately** | **no evidence** | **omitted** | **evidence** | 5 | **expected** | 3 | **observed**

Part C The value of chi-squared

In the light of the expected values calculated in part A and the comments in part B, consider what you should do to use a chi-squared test on this dataset. State the number of degrees of freedom after making any appropriate alterations.

Calculate the value of the chi-squared statistic for this dataset. Give your answer to 4 sf.

Part D The critical value of chi-squared.

Find the appropriate critical value of chi-squared at the 10% significance level. Give your answer to 4 sf.

Part E Hypothesis test

What do you conclude about whether there is evidence for a relationship between the subject chosen and the handedness of the student at the 10% level of significance?

The calculated value of chi-squared is [ ] the critical value at a significance level of 10%. Therefore at this level we [ ] H_0. There is [ ] evidence that subject choice and handedness are related.

Items:

unrelated **significant** **reject** **no significant** **related** **less than** **do not reject** **greater than**

Created for isaacphysics.org by Julia Riley

Gameboard:

STEM SMART Double Maths 43 - Chi-squared Tests

Chi-squared Test: Biased Coins

Further A University

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.

Part A 32 tosses: chi-squared

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 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 : [ ] : [ ] : 1.

H_1 : The coins are biased so that the frequencies of getting 0, 1, 2 or 3 tails [ ] in the ratio 1 : [ ] : [ ] : 1.

The calculated value of chi-squared is [ ] the critical value. Thus we [ ] H_0 at the 10% level; there is [ ] evidence that the ratios are different from 1 : [ ] : [ ] : 1 and the coins are biased.

Items:

less than	1	**are**	2	**reject**	3	**greater than**	**significant**	**are not**	5	**no significant**	**do not reject**	4

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 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 : [ ] : [ ] : 1.

H_1 : The coins are biased so that the frequencies of getting 0, 1, 2 or 3 tails [ ] in the ratio 1 : [ ] : [ ] : 1.

The calculated value of chi-squared is [ ] the critical value. Thus we [ ] H_0 at the 5% level; there is [ ] evidence that the ratios are different from 1 : [ ] : [ ] : 1 and the coins are biased.

Items:

1	do not reject	significant	are	no significant	are not	4	reject	less than	5	2	3	greater than

---

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 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.

H_0 : The coins are such that the frequencies of getting 0, 1, 2 or 3 tails [ ] in the ratio [ ].

H_1 : The coins are such that the frequencies of getting 0, 1, 2 or 3 tails [ ] in the ratio [ ].

The calculated value of chi-squared is [ ] the critical value. Thus we [ ] H_0 at the 5% level; there is [ ] evidence that the ratios are different from [ ].

Items:

$2:5:4:1$ $2:4:5:1$ **less than** **are** $1:3:3:1$ **significant** **do not reject** **are not** **reject** **more than** **no significant** $2:3:2:1$

Created for isaacphysics.org by Julia Riley

Gameboard:

STEM SMART Double Maths 43 - Chi-squared Tests

Chi-squared Test: Active Galaxies

Further A University

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 chi-squared 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% level whether the data are consistent with the assumption that they come from a Poisson distribution.

H_0 : The galaxy distribution [ ] a Poisson distribution with mean as calculated.

H_1 : The galaxy distribution [ ] a Poisson 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 [ ] H_0. There is [ ] evidence that the data do not fit a Poisson distribution with mean as calculated.

Items:

fits **no significant** **less than** **significant** **do not reject** **reject** **greater than** **does not fit**

Created for isaacphysics.org by Julia Riley

Gameboard:

STEM SMART Double Maths 43 - Chi-squared Tests

Chi-squared Test: Measurements

Further A University

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

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, O_i	45	106	122	198	206	205	158	77	63
Expected frequency, E_i	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 A Expected frequencies

Deduce the missing values in the table.

Value	2	4	6	8	10	12	14	16	18
O_i	45	106	122	198	206	205	158	77	63
E_i	56		147		202			99	

Part B The value of chi-squared

Assuming the integer expected frequencies deduced, calculate the chi-squared statistic for the given data. Give your answer to 3 sf.

Part C 5% significance level

Find the critical value for chi-squared using a 5% significance level. Give your answer to 3 sf.

What can you conclude about whether the data follow the proposed distribution at the 5% significance level?

H_0 : The distribution [ ] the proposed model.

H_1 : The distribution [ ] the proposed model.

The calculated value of chi-squared is [ ] the critical value at a significance level of 5%. Therefore at this level we [ ] H_0. There is [ ] evidence that the proposed model does not fit the data.

Items:

does not fit **significant** **reject** **fits** **do not reject** **less than** **no significant** **greater than**

Part D 2.5% significance level

Find the critical value for chi-squared using a 2.5% significance level. Give your answer to 3 sf.

What can you conclude about whether the data follow the proposed distribution at the 2.5% significance level?

H_0 : The distribution [ ] the proposed model.

H_1 : The distribution [ ] the proposed model.

The calculated value of chi-squared is [ ] the critical value at a significance level of 2.5%. Therefore at this level we [ ] H_0. There is [ ] evidence that the proposed model does not fit the data.

Items:

less than | **do not reject** | **no significant** | **significant** | **reject** | **greater than** | **does not fit** | **fits**

Created for isaacphysics.org by Julia Riley

Gameboard:

STEM SMART Double Maths 43 - Chi-squared Tests

Chi-squared Test: Breaking Chains

Further A University

Chains are manufactured and sold in boxes, each of which contains a number of chains. A purchaser requires that the chains should be able to withstand a given force. The manufacturer claims that 60% of the chains will be able to withstand this force.

Fifty boxes of chains are tested. Chains are taken out of a box and tested until one is drawn out that breaks. The observed frequency distribution for the number of chains taken out until one breaks is given in the following table.

Number tested	1	2	3	4	5	6
Frequency	24	17	4	3	1	1

The manufacturer's claim is to be tested at the 10% significance level using a chi-squared test. Answer the following.

Part A Probability distribution

Which of the following is the appropriate probability distribution to use when carrying out the test?

- Geometric distribution
- Normal distribution
- Poisson distribution
- Binomial distribution

Part B Assume manufacturer's claim: chi-squared

The manufacturer's claim that 60% of the chains can withstand the given force is assumed. Using the appropriate distribution, deduce the probability for each outcome and obtain the associated expected frequencies.

Calculate the chi-squared statistic on this assumption. Give your answer to 4 s.f.

Part C Assume manufacturer's claim: hypothesis test

Carry out a chi-squared test to determine at the 10% level whether the model supports the manufacturer's claim.

Find the appropriate critical value of chi-squared. Give your answer to 4 s.f.

Using this critical value and the chi-squared statistic found in part B, what do you conclude about whether the model supports the manufacturer's claim?

H_0 : The data [ ] a [ ] distribution consistent with the manufacturer's claim that 60% of the chains will be able to withstand the given force.

H_1 : The data [ ] a [ ] distribution consistent with the manufacturer's claim that 60% of the chains will be able to withstand the given force.

The calculated value of chi-squared is [ ] the critical value at a significance level of 10%. Therefore at this level we [ ] H_0. There is [ ] evidence that the data do not fit a [ ] distribution consistent with the manufacturer's claim that 60% of the chains will be able to withstand the given force.

Items:

geometric **greater than** **fit** **do not fit** **normal** **do not reject** **binomial** **less than** **accept** **reject** **significant** **no significant** **Poisson**

Part D New estimate of proportion breaking

On the assumption that the data follow the expected probability distribution find a better estimate of the proportion of chains that are expected to withstand the given force. Give your answer as an exact fraction.

Part E New estimate: chi-squared

Using the appropriate distribution and the new estimate of the probability of obtaining a chain that will withstand the given force, deduce the probability for each outcome and obtain the associated expected frequencies.

Calculate the chi-squared statistic on this assumption. Give your answer to 4 s.f.

Part F New estimate: hypothesis test

Find the appropriate critical value of chi-squared. Give your answer to 4 s.f.

What can you conclude about whether the data follow the proposed distribution at the 10% significance level?

H_0 : The data [ ] a [ ] distribution consistent with [ ] of the chains being able to withstand the given force.

H_1 : The data [ ] a [ ] distribution consistent with [ ] of the chains being able to withstand the given force.

The calculated value of chi-squared is [ ] the critical value at a significance level of 10%. Therefore at this level we [ ] H_0. There is [ ] evidence that the data do not fit a [ ] distribution consistent with [ ] of the chains being able to withstand the given force.

Items:

geometric | 46% | **reject** | **Poisson** | 54% | **do not fit** | **accept** | **greater than** | **binomial** | **fit** | **no significant** | 60% | **normal** | 45% | **do not reject** | **less than** | **significant**

Created for isaacphysics.org by Julia Riley