Session 10: Probability

Objectives

By the end of this session, you will be able to

- State that **mathematical probability** is recorded as a number between 0 and 1
- Distinguish between **experimental** probability and **configurational** probability
- Use **symmetry properties** of common shapes to estimate the probability of simple events
- **Conduct an experiment** and compare the results with a prediction of the frequencies of the possible outcomes
- Recognise situations when **adding** and **multiplying** probabilities is appropriate (AND and OR rules)
- Apply probability concepts to contexts including an **approximate model** of DNA matching

The Ancestors

The current world population is estimated to be just over 6 500 000 or 6.5×10^9 people. In 1 000 AD, the world population was estimated to be about 150 000 000 people.

The average time in years between generations is 25 years approximately. You have two parents, they in turn had two parents each, so you had 4 ancestors, three generations ago you had 8 ancestors and so on.

40 generations, or 1 000 years ago, how many ancestors did you have? 2^{40} or 1.1×10^{12} in that generation! The conclusion is that we must all have common ancestors at some point in the past. If this strikes you as blindingly obvious, then read some of the ‘expert evidence’ to be presented later. First, we need to clarify the concept of *probability* or likelihood.

Probability: statistical and configurational

In Maths the chance of something happening is assigned a probability that is expressed as a fraction between 0 and 1. If the favourite in a horse race attracts odds of 5 to 4, then the bookies are saying that there is a 5/9 chance of the horse winning. If you toss a coin, there is an equal chance that the coin will fall showing heads or tails, and we say that the probability of heads is ½ or 0.5. An event with zero probability will never happen, and an event with probability 1 is a certainty. Most events in the real world can be assigned a probability in between 0 and 1.

Your turn 10.1: draw arrows on the scale above showing your best guess of the *probabilities* of the following events…

1. Tony Blair being re-elected as prime minister
2. George Galloway taking part in 'I'm a celebrity, get me out of here'
3. Rolling a fair dice and getting a 5
4. Shuffling a standard 52 card pack of cards and drawing an ace
5. England winning the world cup

You might want to compare results with other members of the class. I'm prepared to guess that the results for items 1, 2, and 5 might show more variation than the probabilities for events 3 and 4.

Events 3 and 4 can be assigned probabilities based on their *configuration* or symmetry properties. A fair dice has 6 sides and each side is equally likely to appear when you roll the dice. A well-shuffled pack gives an equal chance of drawing any particular card. This use of symmetry is a very powerful part of chemistry – you can predict the rates of some reactions from a measure of the probability of a certain kind of collision between atoms or molecules. In general, the probability formula can be applied…

$$\text{probability} = \frac{\text{number of successful outcomes}}{\text{total number of outcomes}}$$

So, for example 4 above, there are 4 aces and 52 cards, and we have

$$\text{probability(ace)} = \frac{4}{52} = \frac{1}{13}$$

For situations where you have no knowledge of the symmetry or configuration, you are forced to rely on history and statistics – the foundation of the insurance industry. You can estimate the future probability of a suitably well-defined event by using a modified version of the probability formula…

$$\text{probability} = \frac{\text{frequency of predicted outcome}}{\text{total number of trials}}$$

So if you surveyed the details of 1 000 car crashes, and found that 175 of the cars driven by the driver at fault were red, you might be tempted to state that

$$\text{probability (red car)} = \frac{175}{1\,000} = \frac{7}{40} \text{ or } 17.5\%$$

Insurance companies do make discounts on car insurance to district nurses and teachers based on the statistical analysis of accident figures – 21 year old City dealers who buy red Porsches can expect to pay rather more for car insurance.

Your turn 10.2: calculating probabilities and frequencies

1) You toss a coin 500 times. How many times would you expect the coin to fall showing **heads**?
2) Suppose you **win a game** by drawing a red picture card (King, Queen, Jack). What is the probability of winning the game?
3) What is the probability of seeing an **even number** when rolling a dice?
4) Suppose you were to toss a coin *twice*. The possible outcomes are HH, HT, TH and TT.
   a) Find the probability of **tossing two heads** in a row
   b) Find the probability of ending up with a **tail and a head** but in *any order*
   c) Find the probability of getting *at least* one head
5) Suppose the probability of the train being early or on time is 80%. What is the probability of the train being late or cancelled?
6) A bag has 3 blue balls, 5 red balls and two yellow balls hidden inside.
   a) What is the probability of picking a **red ball** at random?
   b) What is the probability of picking a ball that is ***not*** **yellow**?
   c) What is the probability of picking a **black ball**?
7) You analyse the weather recorded on the roof of the College for the last 1 000 days. On 16 days in that period, at least one snowflake fell on the roof. Calculate the probability of snow on the College based on this data.
8) Suppose Algernon has tossed two coins 1 000 times and recorded the following results…

HH	207
HT	251
TH	321
TT	221

Do you think that these results are consistent with a **fair unbiased pair of coins**? Can you support your opinion with calculations in any way?

Combining probabilities: AND and OR rules

There are rules for combining the probabilities of different events. Events can be classified as being **independent** or **mutually exclusive**.

Independent events (AND)

Events are **independent** if the outcome of one event has no influence over the outcomes of a second event. Classic textbook examples are the tossing of a coin and the rolling of a dice. The coin can't influence the dice and vice versa. Under these circumstances, you can combine the probabilities by **multiplying** them. The word AND often appears when combining the probabilities of independent events, and so the multiplication of probabilities is referred to in some textbooks as the AND rule.

Example: In a slightly odd game, you toss a coin and roll a dice. You need to score 5 on the dice and the coin must fall tails up to be able to start the game. Calculate the probability of scoring a 5 *and* getting a tail.

P(5 on dice) = $\frac{1}{6}$ p(T on coin) = $\frac{1}{2}$ p(5 and T) = $\frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$

Because you are multiplying fractions, the probability of a combined event is smaller than the probabilities of each event singly. That reflects what you might expect; it is more unusual to get both a 5 *and* a T.

Mutually exclusive events (OR)

Outcomes are **mutually exclusive** if their probabilities add up to one. A classic textbook example is the case of the balls in the bag. You have 5 red balls, 3 green balls and 2 yellow balls in a bag. You shake the bag and invite a friend to pick a ball – the ball *must* have one of the three colours, so the probabilities of picking a red, green or yellow ball *must* add to 1. If you pick a green ball from the bag, then neither of the other outcomes can be achieved at that point – hence the phrase *exclusive*.

As an example, suppose you needed to pull either a red or a green ball from the bag to qualify as a member of some obscure organisation. The probability of this combined outcome is $\frac{5}{10} + \frac{3}{10} = \frac{8}{10} = \frac{4}{5}$.

The probabilities of mutually exclusive outcomes can be *added*, and you often find yourself using the word **OR** to describe situations. This situation is often referred to as the OR rule in maths textbooks.

It is worth mentioning that it is the *outcomes* that are mutually exclusive within a given *event*. For instance, if you pulled a red ball from the bag on a first pick, kept the ball on the table, and then picked a second ball from the bag, the two events (first pick, second pick) are not mutually exclusive although the outcomes of the first event are mutually exclusive, as are the outcomes of the second event.

Dangerous assumptions

Some events are *not* independent, and some outcomes may *not* be fully mutually exclusive. If incorrect assumptions are made about a situation, the calculated combined probability can be completely wrong. Expert witnesses must be *very* careful about using arguments based on combined probabilities, as the case of Professor Sir Roy Meadow illustrates...


BBC NEWS

UK version | International version | About the versions | Low grap

Last Updated: Tuesday, 21 June, 2005, 16:40 GMT 17:40 UK

E-mail this to a friend | Printable version

Paediatrician's 'flawed evidence'

A doctor gave "misleading and flawed" evidence in the trial of a mother wrongly convicted of murder, the General Medical Council has heard.

Paediatrician Professor Sir Roy Meadow is accused of misusing statistics during the 1999 trial of Sally Clark.

VIDEO Sir Roy's background

Professor Meadow was struck off after a BMA hearing (he is appealing against the BMA's decision at the time of writing) – thus ending a long and distinguished career. He suggested that you could treat two successive cot deaths of children within the same family as independent events, and *squared* the statistical probability of a sudden infant death. He then suggested that the *combined* probability was so low that foul play was the only credible conclusion. The verdict in the trial where this evidence was given was subsequently overturned on appeal, although the appeal was granted for other reasons.

Your turn 10.3: Combined probabilities

1) You toss a coin and throw a dice. What is the probability of seeing a head on the coin and rolling a 4 or a 5 on the dice?

2) Nigel has a 0.4 probability of catching the train in the morning. There is a 0.7 chance probability of rain on any given day. Calculate the probability that it is raining *and* Nigel *misses* the train.

3) Aaron has a bag that contains 20 balls. 10 are black, 3 are white, 5 are orange and the rest are beige

   a) How many **beige balls** are there?
   b) What is the probability of pulling a **black *or* white** ball from the bag?
   c) What is the probability of pulling a ball that is ***not* green** from the bag?
   d) What is the probability of picking a **black ball** on the first pick, then replacing the ball and picking a **white ball** on the second pick?
   e) What is the probability of picking a **black ball** on the first pick, *not* replacing the ball, and then picking **another black ball** on the second pick?

4) Calculate the probability of rolling a dice three times and **scoring a 6** on each roll.

5) Calculate the probability of getting **no heads** at all when you toss a coin 4 times.

6) You toss a coin 3 times. What is the chance of getting **at least one head**? Find the *simplest* way of calculating your answer

Simulation experiment

If you have two different coins (say a 2p or a 10p) you can model a simple inheritance experiment. If you work as a group and pool your results, you can get quite good statistics.

Toss your coins 10 times, and record the faces that each coin produces...like my example below...

	1	2	3	4	5	6	7	8	9	10
2p	H	H	T	T	H	T	H	T	T	H
10p	T	H	H	T	T	T	H	H	T	T

Here is a blank table for your results...

	1	2	3	4	5	6	7	8	9	10
2p										
10p										

Then you need to tally each combination of outcomes: 2p always first, 10p always second...

Outcome	MY Frequency	Your Frequency
HH	2	
HT	3	
TH	2	
TT	3	

Record your results here... use the pooled frequency column to record the results of the whole class

Outcome	Your Freq	Group 2	Group 3	Group 4	Group 5	Group 6	Total
HH							
HT							
TH							
TT							

Your turn 10.4: Compare the frequencies for your data, the data from the other groups and for the pooled data with what we expect with is 15:15:15:15 for equal outcomes. Describe the findings in sentences. Are the coins fair? How many of the group results were much further from 15:15:15:15 than expected?

Write your sentences here...

DNA profiling: a *simple* model

There are about 2.5 million DNA profiles held in the UK DNA database – one of the largest in the world. A separate database stores the profiles of Forensic Science service employees and some Police personnel for elimination purposes. When a sample of DNA found at the scene of a crime is checked against the main database, there is *some* chance of a 'false match', i.e. a match that suggests the DNA in the crime scene sample originated from an individual in the database.

Any two people picked at random from the UK population will have a large amount of their DNA in common – recall the back of the envelope calculation we started this session with – any two of us will have a common ancestor at *some* time in the past. To distinguish people by their DNA, "Genetic fingerprinting exploits highly variable repeating sequences called microsatellites. Two unrelated humans will be likely to have different numbers of microsatellites at a given locus. By using the PCR technique to detect the number of repeats at several locii, it is possible to establish a match that is extremely unlikely to have arisen by coincidence." –*Wikipedia* article on Genetic Fingerprinting, 2004.

The very simple model explained below is an attempt to answer the question 'how unlikely is a coincidental match'?

The number of repeats for a sequence at a given locii needs to be established experimentally for each locii, but a rough estimate is that there probability of 0.25 of a match between two 'unrelated' people at any given locus – we shall call this P(match) below.
By carefully choosing different locii, it is possible to regard the probabilities of two samples matching at different locii as *independent events* – this is a *critical* assumption.

The probability of two samples matching at *two* locii is $0.25 \times 0.25 = 0.0625$ assuming independence. When we need to communicate small probabilities to non-expert people (e.g. the Jury) it is usual to quote probabilities as 'one in xxx', so a probability of 0.0625 becomes 'a 1 in 16 chance' of two unrelated people matching at two locii. The reciprocal button on your scientific calculator will help.

Your Turn 10.5: DNA matching

1) Calculate the probability of 6 locii matching in two 'unrelated' samples assuming P(match) is 0.25 for comparisons using any given locus– express your answer as a 'one in XXX' chance.
2) Calcuate the probability of 10 locii producing a match between two 'unrelated' samples – again use the 'one in' language to communicate the results
3) Repeat questions 1 and 2 using a probability P(match) = 0.5 for matching at a locus for two 'unrelated' people – write a sentence comparing your results for the two P(match) values with 6 and 10 locii

4) Copy and complete the table below… put the 'chances' in the cells of the table I have completed the 5 locii column as a guide.

	Number of locii			
P(match)	5	10	15	20
0.2	3125			
0.25	1025			
0.5	32			

Harder questions (stretch)

5) Suppose the P(match) was 0.25 for any two 'unrelated' samples and a given locii. Can you 'work backwards' and estimate roughly *how many locii* would need to be compared to reduce the chance of a complete match to less than 1 in 50 million? Hint: *logs*

6) Can you use a spreadsheet to build a *formula* that provides the answer to question 5 accepting the probability P(match) and giving a rough answer? Use your spreadsheet to write a sentence explaining how *sensitive* the dependence of the chance of complete matching of two 'unrelated' samples is on the P(match) chosen.

Summary of session 10

Complete the gaps in this paragraph…

A mathematical probability is always measured as a ________ between 0 and 1. A probability of 1 means ___________. A probability of 0 means ____________. Probablilities can be worked out by analysing the results of many experiments or trials – such probabilities are called ___________ probabilities. You can estimate the probability of an event from the symmetry of the situation sometimes, probabilities estimated using symmetry are called ______________ probabilities. The probabilities of independent events can be _______________. The probabilities of mutually exclusive outcomes can be _______________.

Then write your *own* sentences explaining each of the following terms.

An event	
An outcome of an event	
Mutually exclusive events	
Independent events	

Your turn answers

10.1 3) $\frac{1}{6}$, 4) $\frac{4}{52}=\frac{1}{13}$ I would not *dream* of arguing against your answers to 1,2, or 5.

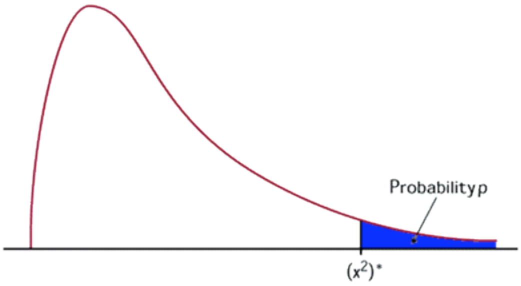
10.2 1) 250, 2) $\frac{12}{52}=\frac{3}{13}$, 3) $\frac{3}{6}=\frac{1}{2}$, 4) a) $\frac{1}{4}$ b) $\frac{1}{2}$ c) $\frac{3}{4}$, 5) 20%, 6) $\frac{5}{10}=\frac{1}{2}$ 7) $\frac{16}{1000}=\frac{2}{125}\approx 1.6\%$ 8) There is a way of quantifying how far from the calculated probabilities a result is: see next session!

10.3 1) $\frac{2}{6}\times\frac{1}{2}=\frac{1}{6}$, 2) 0.4 × 0.7 = 0.28, 3) a) 2 beige balls, b) $\frac{13}{20}$, c) 1, d) $\frac{10}{20}\times\frac{3}{20}=\frac{3}{40}$, e) $\frac{10}{20}\times\frac{9}{19}=\frac{9}{38}$, 4) $\frac{1}{6}\times\frac{1}{6}\times\frac{1}{6}=\frac{1}{216}$, 5) $\frac{1}{16}$, 6) $1-\frac{1}{8}=\frac{7}{8}$

10.4 Depends on your results! We will re-visit this data set in Session 11

10.5 1) 1 in 4096, 2) 1 in 1 048 576, 3) 1 in 64 and 1 in 1024 – a sentence might include a comment on the fact that the chances of a complete match increase for the higher value of P(match) and if you are *really* on the ball, the fact that the chance for the 10 locii changes by a much larger factor than for the 6 locii – so the effect of small changes in P(match) is greater for larger locii tests, 4) $\frac{\log(50000000)}{\log(\frac{1}{0.25})}=12.7...\approx 13$ locii, 5) you will find that the number of locii depends very strongly on the value of P(match) assumed.

Session 11: Scenario 4



By the end of the session, you will be able to

- Calculate the expected values for various experiments
- Calculate the χ^2 statistic for two different formats of table (2 x 1 and 1 x n)
- Describe a null hypothesis appropriate for the fruit fly breeding experiment
- Perform a significance test at the 5% level on the null hypothesis chosen for a fruit fly dataset. Evaluate the results.
- **Advanced**: Calculate the Yates continuity correction for the case of a 2 × 1 table
- **Advanced**: pool all the datasets and re-perform the test, explain how the results change
- **Advanced**: re-perform a test with expected observations based on a non-Mendelian 2:1 ratio. Comment on the extent to which the data can discriminate between the 2:1 or the 3:1 ratio

Overview

This session introduces the χ^2 statistic – a way of calculating how far your results are from the 'predicted' or 'expected' results. If the results of the experiment are 'too far' from the expected values, then we have to look at the assumptions we used to calculate the predicted values in the first place. Later in the session, we will use tables of 'critical' values of the χ^2 statistic for various probability levels – the procedure is very similar to the one used in Scenario 2 and is another example of a *statistical test.* Before we can do any of this, you need to know how to estimate the 'expected values' from a set of assumptions.

Calculating the expected values

Suppose you toss a pair of identical 2p coins a large number of times. You would expect the numbers of each outcome (HH, HT, TH, TT) to be the same, however, you can't *distinguish* the HT and TH outcomes as you can't tell which 2p coin was which. You therefore combine the HT and TH outcomes and call the combined outcome 'one of each'. You might expect the three outcomes HH, One of Each, TT to be in the ratio 1:2:1, i.e. twice as many 'One of Each' results as TT or HH.

To calculate the expected value for a given number of trials, just find the probability of each outcome (divide each ratio term in turn by the total of the ratio terms), then *multiply* by the number of trials… look at the example below.

Suppose you tossed the pair of new mint 2p coins 100 times. The expected results might look like this…

Outcome	**Method**	**Expected value**
HH	• Add the ratio terms to find the number of parts • Divide the ratio term by the number of parts to find the 'probability' of the outcome • Multiply by the number of trials	$\frac{1}{(1+2+1)}\times 100 = 25$
One of Each		$\frac{2}{(1+2+1)}\times 100 = 50$
TT		$\frac{1}{(1+2+1)}\times 100 = 25$

Your turn 11.1: calculate expected values

1) Suppose you have an experiment where you think the results might be in the ratio 9:3:3:1, for instance crossing double dominant (yellow flower (A) and round seeds (B)) and double recessive (white flower (a) and wrinkled seeds (b)) pea plants. The offspring will include Yellow flower/round seeds, White flower/round seeds. Suppose you conduct 160 trials. Calculate the expected values for each outcome – write your calculations in the table below.

Phenotype	Genotypes (equally likely)	Ratio	Expected value
Yellow flower, round seed	AABB, AABb, AAbB AaBB, AaBb, AabB aABB, aABb, aAbB	9	
Yellow flower, wrinkled seed	AAbb, Aabb, aAbb	3	
White flower, round seed	aaBB, aaBb, aabB	3	
White flower, wrinkled seed	aabb	1	

2) You can have fractional expected values: Suppose you toss two coins (again) and track the two outcomes 'No heads' and 'at least one head'.
   a) Are these outcomes mutually exclusive?
   b) Calculate the ratio of 'No Heads' to 'At least one head' outcomes
   c) Calculate the expected values for 10 throws of the pair
   d) Calculate the expected values for 30 tosses of the pair

Use a piece of A4 paper to draw up a table (use a ruler!) to show your calculations.

Calculating the χ^2 statistic

Once you have your 'expected values' (and have checked the outcomes are mutually exclusive so that the probabilities add to one) and your 'observed values' from the actual trial, you can calculate the χ^2 statistic for the results. The χ^2 statistic is simply a number that measures how 'far away' the observed results are from the expected results.

O is always used for the Observed results, and *E* is always used for the Expected results, and the formula for the χ^2 statistic is

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

As a concrete example, suppose we took the coin tossing data from Your Turn 10.2, question 8. Algernon has tossed a pair of different coins 1 000 times, and recorded the outcomes below

HH	207
HT	251
TH	321
TT	221

Algernon wants to know if these results are consistent with unbiased coins. Calculating the χ^2 statistic will at least enable us to give a measure for the 'distance' between the observed and expected results. The expected values in this case take a very simple form since all four outcomes are equally likely (and mutually exclusive – as they must be for a valid χ^2 calculation) – just 250 of each outcome.

The classic layout for the calculation is to use a table like the one below – you can find *plenty* of worked examples on the Web.

	Outcomes				
	HH	HT	TH	TT	Totals
Observed O	207	251	321	221	1 000
Expected E	250	250	250	250	1 000
$O-E$	– 43	1	71	– 29	0 (check)
$(O-E)^2$	1849	1	5041	841	(not needed)
$\frac{(O-E)^2}{E}$	7.396	0.004	20.164	3.364	$\chi^2 = 30.928$

The total of the $\frac{(O-E)^2}{E}$ terms gives the value of χ^2 in this case 30.928. As we will see in the next section, this is a *huge* value of χ^2 and the coins are *very* likely biased!

Your turn 11.2: Calculating the χ^2 statistic

1) Suppose you toss a pair of identical 2p coins 30 times and tally the three outcomes HH, 'One of Each' and TT. Suppose you get results like this…

Outcome	Frequency
HH	8
One of Each	13
TT	9

Calculate the χ^2 statistic for this data – use the table layout below…

	Outcomes			
	HH	One of Each	TT	Totals
E	7.5	15	7.5	30
O	8	13	9	30
$O–E$				
$(O-E)^2$				
$\frac{(O-E)^2}{E}$				$\chi^2 =$

Suppose you tossed two coins 25 times and found that you had 9 trials with no heads at all. Calculate the expected values for the two mutually exclusive outcomes in this experiment ('no heads' and 'at least one head'), and then calculate the χ^2 statistic for this data.

Space for calculations and table – the latter *drawn with a ruler*!

Performing a χ^2 test

Having calculated a statistic that tells us 'how far' the observed results are from the expected results, we now need a criterion for deciding 'how far is too far'. You can then decide if your results confirm or refute your hypothesis. The hypothesis you are testing using a χ^2 test is always the same:

Null Hypothesis: there is ***<u>no difference</u>*** between the Observed and Expected results.

Statistical tables of χ^2 values show the probability that a given value of χ^2 would occur by chance. Typical probability levels are 0.05 (5%) and 0.01 (1%).

- If the χ^2 value you calculate is higher than the χ^2 value in the table for your chosen probability, then you **reject** the null hypothesis of no difference between your observed values and the expected values
- If the calculated χ^2 value is lower than the χ^2 value in the table, then you accept the null hypothesis of no difference and sigh a deep sigh of relief

There is one extra complication: **degrees of freedom**.

Degrees of freedom

The 'degrees of freedom' in an experiment is defined the number of *independent* random variables in the experiment. As you know the total number of trials, the 'degrees of freedom' is *usually* one less than the number of outcomes.

Recall that a χ^2 test should only be used for frequencies or counts (*not* percentage data or proportions) within mutually exclusive categories (so the probabilities of all outcomes will always add to 1). Suppose you have an experiment with four mutually exclusive outcomes: as you know the total number of trials, only three of the outcomes can vary randomly – the frequency of one of the outcomes is fixed when the other three are known. This situation therefore only has 3 degrees of freedom.

If you encounter data that has a 'two dimensional' table of outcomes in the future then the degrees of freedom will be one less than the number of rows multiplied by one less than the number of columns: $(R-1) \times (C-1)$.

Worked example

Recall the example based on the tossing of a pair of coins 1 000 times. The χ^2 value was calculated as 30.928 and there were 4 outcomes (HH, HT, TH, TT). The layout for the test at a 5% probability is given below (I have duplicated the table to show a sample layout for your own work)

	Outcomes				
	HH	HT	TH	TT	Totals
Observed O	207	251	321	221	1 000
Expected E	250	250	250	250	1 000
$O - E$	– 43	1	71	– 29	0 (check)
$(O - E)^2$	1849	1	5041	841	(not needed)
$\frac{(O-E)^2}{E}$	7.396	0.004	20.164	3.364	$\chi^2 = 30.928$

Null hypothesis: there is no difference between the Observed and Expected frequencies of the outcomes

Degrees of freedom: 4 – 1 = 3 (4 outcomes in a single 'row' so one less independent random variables).

Use a probability level of 5%

- Entering a table of χ^2 values at 5% (0.05) for 3 degrees of freedom gives χ^2_{table} = 7.82.
- The calculated $\chi^2_{experimental}$ = 30.928 and that is larger than the table value for 5% probability.
- We are therefore forced to reject our null hypothesis of no difference, and we must conclude that there is a statistically significant difference between the observed outcome and what we would expect from a pair of unbiased coins.

We accept that there is a 1 in 20 chance that actually, the coins are fair, and that the frequencies we observed were a fluke or freak outcome.

If we used a probability level of 1% (0.01) and 3 degrees of freedom,

- Entering the table at 1% and df=3 gives χ^2_{table} = 11.35
- The $\chi^2_{experimental}$ = 30.928 is still much larger
- We reject the null hypothesis of no difference again

Now there is much less than a 1 in 100 chance that the coins are really fair and that the frequencies we saw were a sport of nature.

Your turn 11.3: perform a χ^2 test

Now for a practice run before you tackle Scenario 4.

Dusk is falling at the Station on a crisp February evening, and Rakesh and Richard are having a discussion about probability. Rakesh suggests that there was a time in history when it was accepted that if you toss two identical coins – totally indistinguishable – the outcomes HT and TH were regarded as the *same*, and counted as *one* equally likely outcome.

Richard challenges this assertion and points out that a simple trial would have decided the issue, even in 16^{th} century France. Even if the coins were indistinguishable, the 'One Of Each' outcome would appear around 50% of the time and would have made the error in the argument apparent.

Rakesh counters with the view that the statistics of trials with a small number of tosses would not allow the early mathematicians to reject the hypothesis of equal probability for the outcomes HH, One of Each and TT.

Quickly, a pair of identical newly minted coins are found, and a trial begun. The coins are tossed 30 times. The activity attracts the attention of other people in the Station and a small audience gathers…

	Results
HH	10
One of Each	13
TT	7

1) Calculate the **expected values** for this experiment based on the assumption that there is a 1:2:1 ratio between the outcomes HH, One of Each and TT
2) Perform a χ^2 test on the data at the 5% probability level for the expected values corresponding to the 1:2:1 ratio
3) Now **re-calculate** the expected values on the basis of the 1:1:1 ratio and re-perform the χ^2 test at the 5% level
4) Can you **reject the null hypothesis** of no difference between the expected and observed values for either of the ratios?
5) **Compare** the two values of χ^2 obtained in the two tests – which ratio gives the largest χ^2 value and is thus 'further away' in some sense from the observed data
6) What would you recommend that Rakesh and Richard do to decide the argument?

Paragraph about F1 Mendelian cross. AA and aa give AA, aA, Aa and aa offspring. Mendel faced a situation where many gardeners and botanists regarded the aA and Aa as identical, and thus calculated the ratio of aa to other genotypes as 1:2 instead of the correct 1:3. The situation was complicated as gene linkage was not understood - Mendel used genes with no linkage but others used genes that had strong linkages to other traits.

Yates continuity correction

For results that have 1 degree of freedom, the χ^2 test may no longer be the correct test to apply. It is possible to apply something called an 'exact test' to the results – suppose you had 60 flies and were expecting to see vestigial winged flies one quarter of the time, and normal winged flies in 45 cases. An exact test would calculate a probability associated with every possible outcome (1 and 59, 2 and 58, 3 and 57 and so on) and then you would simply look up the *probability* of your results and decide on acceptance or rejection of the null hypothesis accordingly. The calculations needed are quite sophisticated and beyond the scope of this unit.

For situations where you have more than around 40 results in the table, and no cell has less than 5 counts in it, a correction to the χ^2 test was devised by Yates – a statistician working in the early 20th century. The correction is simply to take 0.5 off the *absolute* value of O – E before squaring the difference. The new formula is

$$\chi = \sum \frac{(|O-E|-0.5)^2}{E}$$

The vertical bar notation $|O-E|$ means take the *absolute value* of the difference, ignoring the sign. Suppose O was 15 and E was 18, you would then *ignore* the sign of the difference and calculate 3 – 0.5 and square the result.

Worked example

Suppose we had 17 flies with vestigial wings and 43 flies with normal wings. As the vestigial winged flies correspond to two *recessive* genes (ww) and all other gene combinations correspond to phenotypes with fully developed wings (wW, Ww and WW) we expect a ratio of 1:3 in the offsrpring. The table below shows the calculation in full for this situation…

	Outcomes		
	V	W	Totals
O	17	43	60
E	15	45	60
$O-E$	2	−2	0
$\|O-E\|-0.5$	1.5	1.5	
$(\|O-E\|-0.5)^2$	2.25	2.25	
$\frac{(\|O-E\|-0.5)^2}{E}$	0.15	0.05	$\chi^2 = 0.2$

The χ^2 value without the Yates correction will be $\frac{4}{15}+\frac{4}{45}\approx 0.35$ so the correction makes quite a difference at small χ^2 values. A difference of 0.15 or so for χ^2 values near 3.6 is less important.

Your turn 11.4: Calculate the χ^2 value applying the Yates correction for the following fly data – 20 vestigial winged flies and 40 flies with fully formed wings.

Summary of session 11

- There is a number called the χ^2 statistic that can measure 'how far' a set of observed count data is from the values expected based on a theory you are trying to test
- The χ^2 statistic can only be used with frequency or count data or with actual values – you can't use the χ^2 statistic with percentages or ratios directly
- You need a total frequency of at least 40 and each cell in your Observed data table must have a frequency higher than 5 to use the χ^2 test reliably
- To calculate the χ^2 statistic, simply find each difference between the expected and the observed count for each cell in your table of results, square the difference and divide each by the expected value. Adding the various contributions gives you the χ^2 statistic for that data and expected values
- The χ^2 test always tests the null hypothesis that: "there is no difference between the expected and observed results"
- You perform the χ^2 test by calculating the χ^2 statistic for your data, and then looking up a critical value of χ^2 in the tables for a probability level (usually 0.05 or 0.01) and a number of 'degrees of freedom'
- The degrees of freedom for the table of results is one less than the number of rows multiplied by one less than the number of columns except when there is only one row – then the DF is simply one less than the number of rows
- If you have a 2 × 2 table or a single row with just two numbers (i.e. data sets with one degree of freedom) then you can apply Yates' continuity correction. Just take the positive difference between the Observed and Expected counts for each corresponding cell, and subtract 0.5. Then square the result, and divide by the Expected value to calculate the contribution towards the χ^2 statistic. Adding both contributions gives the total corrected χ^2 value – which is always reduced slightly compared to not applying the correction.

Your Turn answers

11.1 1) 90, 30, 30, 10 flies 2) a) Yes b) 1:3 c) 2.5, 7.5 d) 7.5, 22.5

11.2 $\chi^2 = \frac{0.25}{7.5} + \frac{4}{15} + \frac{2.25}{7.5} = 0.6$

11.3 1) 7.5, 15, 7.5 and 10, 10, 10 for the two ratios 2) $\chi^2 = 1.13$ 3) $\chi^2 = 1.80$
4) No, both χ^2 values are well below the critical value from tables for 2 DF and p = 0.05 (5.99) so we must accept the null hypothesis of no difference in both cases! The implication is that Richard and Rakesh did not toss the coins enough times to be able to distinguish between the two theories. Rakesh is ahead on points at present. 5) The 1:1:1 ratio theory produces a χ^2 value that is a little bit higher than the χ^2 value produced by the 1:2:1 theory, so you might argue that the latter fits the data better. 6) Do more trials!

11.4 $\chi^2 = \frac{(|20-15|-0.5)^2}{15} + \frac{(|40-45|-0.5)^2}{45} = 1.35 + 0.45 = 1.8$ and this value is well below the 3.84 critical value from tables corresponding to for p = 0.05 and 1 df.

Scenario 4

Objectives

In this session, you will

- Action plan a meta-analysis of the results of a series of fly breeding experiments
- Carry out your meta-analysis
- Discuss the results
- Recommend pooling sets of results to improve the statistical power of the experiment
- Evaluate the reliability of your results

The results

The table below shows the results of a number of breeding experiments with fruit flies. Each data set (A, B and so on) was collected by students counting their own set of fruit flies. The total numbers of flies in each set depends on the success in breeding the flies, and in keeping them captive long enough to count!

Ref	Winged	Vestigial
A	40	23
B	58	18
C	53	24
D	52	19
E	52	21
F	49	16

Action planning the analysis

Will you pool results or test each data set separately and then pool results?	
What null hypothesis will you use?	
What probability level will you use?	
What ratio(s) will you use to calculate the expected values?	

Post analysis points

Did you decide to include or exclude data set A from the pooled data set?	
Is the pooled data set large enough to be able to reject the null hypothesis for expected values calculated on the basis of a 1:2 ratio?	
Look at the critical value of the χ^2 statistic for *95%* and *df* = 1. Are any of your calculated χ^2 values less than or equal to this value?	
Can you claim to have established the 1:3 ratio on the basis of this data set?	

Having completed scenario 4, you need to pull everything together in the next session.

Appendix 1: Statistical tables

Table of percentage points of the χ^2 distribution

Degrees of Freedom	Probability, *p*				
	0.99	**0.95**	**0.05**	**0.01**	**0.001**
1	0.000	0.004	3.84	6.64	10.83
2	0.020	0.103	5.99	9.21	13.82
3	0.115	0.352	7.82	11.35	16.27
4	0.297	0.711	9.49	13.28	18.47
5	0.554	1.145	11.07	15.09	20.52
6	0.872	1.635	12.59	16.81	22.46
7	1.239	2.167	14.07	18.48	24.32
8	1.646	2.733	15.51	20.09	26.13
9	2.088	3.325	16.92	21.67	27.88
10	2.558	3.940	18.31	23.21	29.59
11	3.05	4.58	19.68	24.73	31.26
12	3.57	5.23	21.03	26.22	32.91
13	4.11	5.89	22.36	27.69	34.53
14	4.66	6.57	23.69	29.14	36.12
15	5.23	7.26	25.00	30.58	37.70

A significant difference from your null hypothesis (i.e. difference from your expectation) is indicated when your calculated χ^2 value is greater than the tabulated value shown in the 0.05 column of this table (i.e. there is only a 5% probability that your calculated χ^2 value would occur by chance). You can be even more confident if your calculated value exceeds the tabulated values in the 0.01 or 0.001 probability columns.

If your calculated χ^2 value is equal to, or less than, the tabulated value for 0.95 then your results give you no reason to reject the null hypothesis.

In a few special circumstances (though not generally) a calculated χ^2 value lower than the tabulated value in the 0.95 or 0.99 columns provides evidence that your results agree well with a null hypothesis.