re(Solve) MATHS BY INQUIRY Hot Streaks

Y9

Summary of learning goals

- To develop tests for randomness in order to distinguish between random and non-random results. Students use their understanding of randomness to investigate the existence or otherwise of the 'hot streak' phenomenon in basketball.

Australian Curriculum: Mathematics (Year 9)

ACMSP226: Calculate relative frequencies from given or collected data to estimate probabilities of events involving 'and' or 'or'.

Summary of lessons

Who is this sequence for?

- This sequence is for students who are familiar with using spreadsheet programs to analyse data and who are ready to create their own methodologies to explore more complex datasets.

Lesson 1: Pick the Fake Data

Students choose between flipping a coin 50 times and recording the results or making up a fake sequence of 'random' results. They then experiment with different statistical methods to find a strategy for identifying whether other students' results are fake or are truly randomly generated.

Lesson 2: Do Hands Get Hot?

Students learn about the hot hand phenomenon in basketball. Using their findings from Lesson 1: Pick the Fake Data, students explore different methodologies to prove whether the phenomenon exists.

reSolve: Mathematics by Inquiry is funded by the Australian Government and is managed by the Australian Academy of Science.

Reflection on this sequence

Rationale

Randomness is a fundamental, but often misunderstood, concept in statistics. Misunderstanding randomness lies at the heart of the gambler's fallacy, which is the belief that if something happens more frequently than normal during a given period, it will happen less frequently in the future. This sequence looks at the reverse of this in the context of a hot streak in basketball – does success lead to more success?

reSolve mathematics is purposeful

- The sequence deals with the fundamental idea of randomness, which is essential for developing statistical reasoning.
- The sequence uses a very large dataset to address a real-life problem that is of interest to sports statisticians.

reSolve tasks are inclusive and challenging

- Students engage in a common activity of generating and analysing fake or random data as an introduction to how large datasets might be analysed.
- Lesson 2: Do Hands Get Hot? deals informally with the concept of conditional probability.

reSolve classrooms have a knowledge-building culture

- The use of spreadsheets allows students to compare data and make decisions about the randomness, or otherwise, of a given dataset.

Further reading

For discussion on appropriate methodology for testing the existence of the hot hand, see:
Gilovich T, Vallone R and Tversky A, 1985, 'The hot hand in basketball: On the misperception of random sequences', *Cognitive Psychology* 17: 295–314.
Burns BD, 2001, 'The hot hand in basketball: Fallacy or adaptive thinking?', *Proceedings of the Annual Meeting of the Cognitive Science Society*, 23.
Korb K and Stillwell M, 2003, 'The story of the hot hand: Powerful myth or powerless critique?' presented at the International Conference of Cognitive Science.
Miller JB and Sanjurjo A, 2018, 'Surprised by the gambler's and hot hand fallacies? A truth in the law of small numbers', *Econometrica*, 86: 2019–47
Woo E, 2018, *Woo's Wonderful World of Maths*, Pan Macmillan Australia, Sydney. This book discusses sequences of fake or random heads and tails in Chapter 18 Conspiracy theory.

Acknowledgements

The NBA data used in Lesson 2: Do Hands Get Hot? are taken from www.basketball-reference.com.

LESSON 1

reSolve MATHS BY INQUIRY Hot Streaks

Pick the Fake Data

Y9

About this lesson

Students choose between flipping a coin 50 times and recording the results or making up a fake sequence of 'random' results. They then experiment with different statistical methods to find a strategy for identifying whether other students' results are fake or are truly randomly generated.

Australian Curriculum: Mathematics (Year 9)

ACMSP226: Calculate relative frequencies from given or collected data to estimate probabilities of events involving 'and' or 'or'.

Mathematical purpose

- To learn methods of distinguishing fake from real randomly generated results, using a collection of statistical tools.

Learning intention

- To design a method for finding whether a 'random' sequence is truly randomly generated.

Time

Two lessons of approximately 1 hour each.

Vocabulary

- random

Resources

- Student preparation required as outlined in Teacher background information (at least one day in advance)
- reSolve PowerPoint *1a Pick the Fake Data*
- Student Sheet 1 – Making a Macro (optional)

reSolve: Mathematics by Inquiry is funded by the Australian Government and is managed by the Australian Academy of Science.

Teacher background information

Expected coin flip results

This lesson explores what results we would expect from random coin flips. What is the number of times that we would expect to flip the same result in a row (i.e. the expected streak length) for fair coin flips?

Below is an explanation for how the expected streak length can be calculated. It is not taught as part of the lesson and is included here only for teacher interest.

Imagine flipping a coin. For a streak length of 1, the second flip must be different from the first. The probability that the second flip is different from the first is $\frac{1}{2}$. So the probability of a streak length of exactly 1 is $\frac{1}{2}$. We can write this as $\mathrm{P}(L = 1) = \frac{1}{2}$, where L represents the length of a streak.

For a streak length of 2, the second flip must be the same as the first, then the third flip must be different.

P(flip same as previous flip) $= \frac{1}{2}$.

P(flip different from previous flip) $= \frac{1}{2}$.

Since each coin flip is an independent event, we can multiply their probabilities to find the probability of a streak length of exactly 2: $\mathrm{P}(L = 2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

For a streak length of 3, we need same, same, different, so $\mathrm{P}(L = 3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$. And so on.

The expected streak length can be found by multiplying each possible outcome by the probability of that outcome and then adding all values:

$$\mathrm{E}[L] = 1 \times \mathrm{P}(L = 1) + 2 \times \mathrm{P}(L = 2) + 3 \times \mathrm{P}(\mathrm{L} = 3) + 4 \times \mathrm{P}(\mathrm{L} = 4)$$

$$= 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 4 \times \frac{1}{16} + \ldots$$

Multiplying by 2: $2\mathrm{E}(L) = 1 + 1 + \frac{6}{8} + \frac{8}{16} + \ldots$

$$2\mathrm{E}(L) - \mathrm{E}(L) = \mathrm{E}(L) = 1 + (1 - \tfrac{1}{2}) + (\tfrac{6}{8} - \tfrac{2}{4}) + (\tfrac{8}{16} - \tfrac{3}{8}) = 1 + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \ldots = 2$$

Therefore, when flipping a fair coin, the expected streak length is 2.

Student preparation required

At least one day before presenting this lesson, give students the homework of flipping a coin 50 times and recording the results. They can also use Microsoft Excel or a random number generator to generate 50 0s and 1s. Alternatively, they may choose to make up a sequence of 50 'random' results from off the top of their heads.

Students should record their results as a string of 50 Hs and Ts, and secretly record whether their results are made up or truly randomly generated via coin flips or Excel.

These result strings need to be collected and made available to the class so that all students can view the complete set of classroom results. This could be done by having students add their results to a group Google spreadsheet or asking them to submit their results and then collecting them in a single spreadsheet. Each string of 50 results must be numbered or named so that students can identify their string.

Introduction

Explain that in this lesson students will be trying to devise a system to identify whether other students' coin flip sequences are truly randomly generated.

Resources: Show students the three (fake) coin flip sequences on slide 2 of reSolve PowerPoint *1a Pick the Fake Data*.

Ask students: *Do you think these coin flip results are real or fake? What makes you think that?*

Have students make a list of the reasons why they are confident that these sequences are fake. It is important for students to notice how *long* the runs of heads or tails are (referred to below as the 'maximum streak length') and how *often* the sequence switches between heads and tails (which can also be thought of as the 'mean streak length'), as these are key indicators of randomness.

Possible student responses:

- The sequence changes between heads and tails too often; that is, the '**mean** streak length' is too short.
- The sequence doesn't change between heads and tails enough; that is, the '**mean** streak length' is too long.
- There are too many heads in a row or too many tails; that is, the '**maximum** streak length' is too long.
- The proportion of heads is too extreme (or too exact); that is, 'there are too many heads' or 'getting exactly 50% heads or tails makes me suspicious'.

Show students the more complicated example on slide 3: *Is this sequence real or fake?* Guide the class towards the conclusion that they need to collect more information on what real coin flip sequences look like.

Generating random sequences

Have each student generate a sequence of 50 coin flips in Microsoft Excel using the formula "=IF(RAND()>0.5, "H", "T")".

Teacher notes:

- This formula:
  1. Generates a random number between 0 and 1 in each cell.
  2. Checks if the random number is > 0.5.
  3. Replaces numbers over 0.5 with H and numbers under 0.5 with T. This means that each cell has a 50% chance of containing a H and a 50% chance of containing a T.

Ask students: *How can you find the maximum streak length and the mean streak length of your sequence?*

Allow students some time to experiment with ways to calculate these values in Excel. Students might find the following two formulas useful, which are also given on slide 4.

- "=IF(X,Y,Z)": If X is true, then Y; otherwise Z.
  - For example: =IF(A1=A2, "Same", "Different") checks if cell A1 matches cell A2. If they are the same, the formula prints "Same"; otherwise, the formula prints "Different".
- "=COUNTIF(A1:B26,"X")": Counts the number of cells between A1 and B26 that contain X.
  - For example: =COUNTIF(A1:A50, "Same") counts the number of cells between A1 and A50 that contain the word "Same".

Resources: See reSolve Excel Spreadsheet *1b Annotated Coin Flip Analyser* for a possible approach to constructing the spreadsheet.

Encourage students to also calculate any other statistics that they think might be informative/relevant, such as the total number of heads and the total number of tails in the sequence.

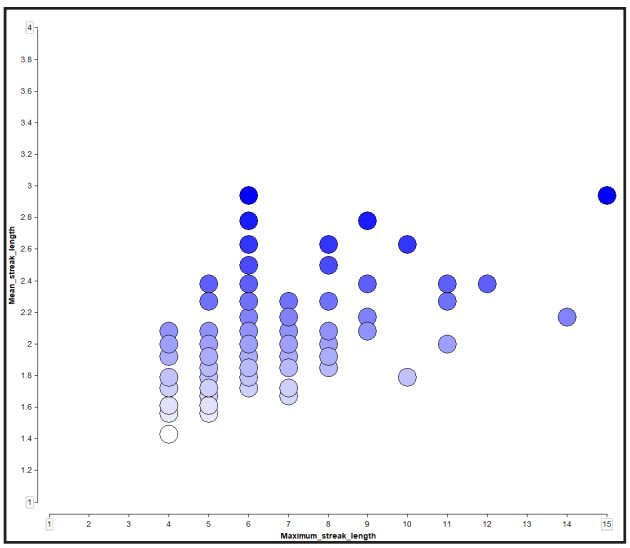
Have students record the mean and maximum streak lengths in a new spreadsheet, as well as any other statistics they have calculated, then use the F9 shortcut or the 'recalculate sheet' button to generate a new random set of coin flips and record the new values. Students should repeat this process 100 times to create a useful dataset.

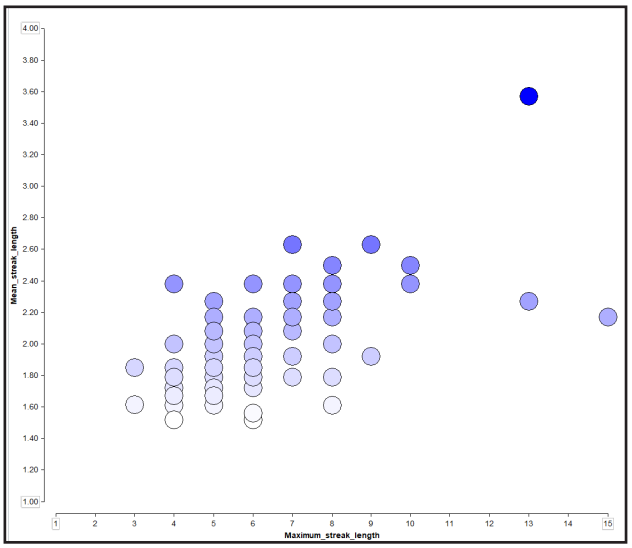
Teacher notes:

- Students might choose to work in groups to create a large enough dataset or to create a macro to do this automatically (see Student Sheet 1 – Making a Macro for instructions in Excel 2010 or 2013).

Analysing sequences

Import these values into CODAP (https://codap.concord.org/releases/latest/static/dg/en/cert/index.html), Tinkerplots or graph them using another program. Below are two examples that use two different sets of 100 randomly generated coin flip sequences (to illustrate the variation that will exist across the class), which graph the maximum streak length (*x*-axis) against the mean streak length (*y*-axis).

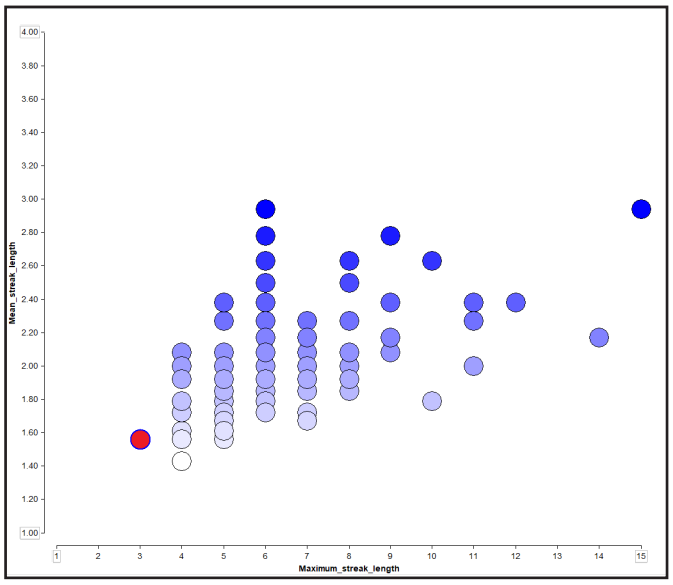




Discuss with students what the points on the graph represent.

Ask: *Where would you find a sequence that has frequent changes between heads and tails? Where would you find a sequence that has a long run of heads or tails?*

Ask students to input the sequence from slide 3 of reSolve PowerPoint *1a Pick the Fake Data* into their spreadsheet, find its maximum and mean streak lengths, and then mark its location on the graph. It is highlighted in red in the graph at right (also shown on slide 5).



Discuss with students: *Does this sequence seem likely to be truly randomly generated or is it fake?*

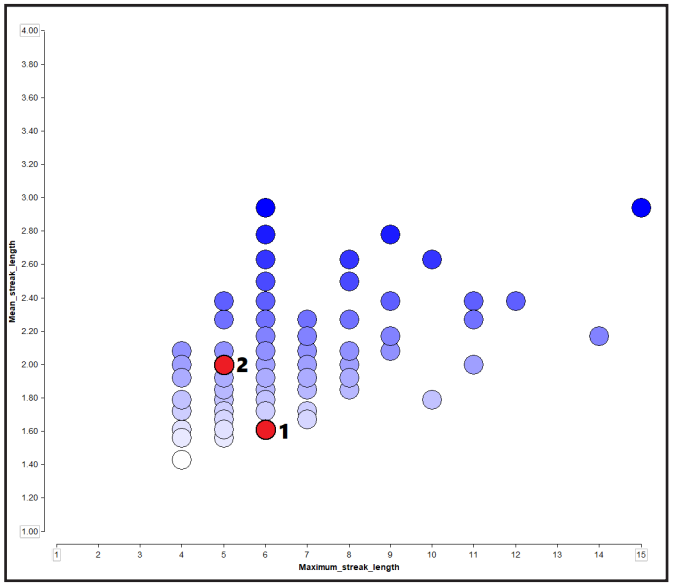
Students should identify that it is likely to be fake, as it is an outlier on the graph.

Ask students to create a simple rule to determine whether a result is likely to be real or fake, which will be refined later. At this stage, the rule might be as simple as: *If the mean streak length is > 1.5 and the maximum streak length is ≤ 4, the result is likely to be real.*

Ask students to test the examples below, which are also given on slide 6. Hint that one sequence is real and one is fake. Ask: *Which is more likely to be the real sequence? Why?*
Reveal the answer and ask students to refine their rules. The graph below is shown on slide 7.

1. HTHHHTTHTHHHHHHTHTHTTHTTHTHHHTTTHTTTHHTTHTHHTHTHTH [fake]
2. THTHHHHTTTTTHTTHHHTHHHTTTHHHTHHHTHHHTHTHTHHTHHHTTT [real]

Test each of the classroom-submitted sequences and determine whether they are real or fake.



Reflection

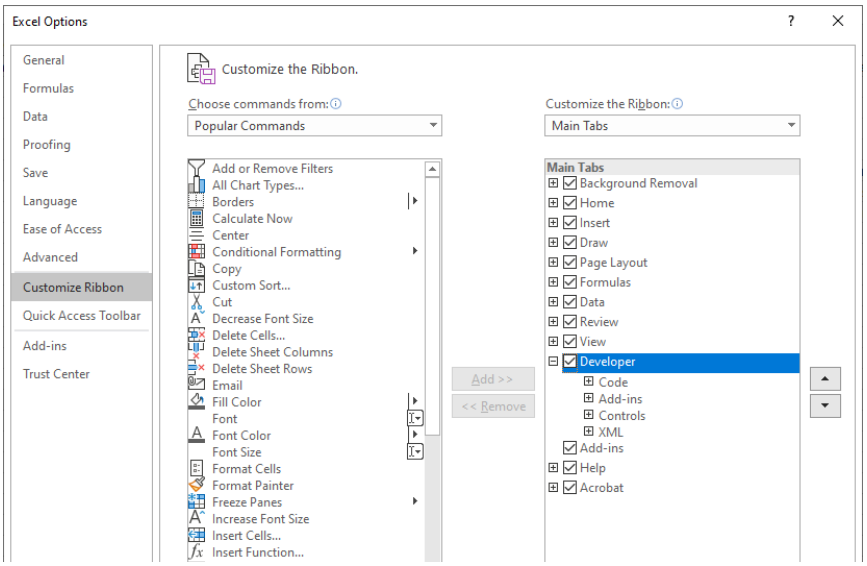
As a class, discuss:

- Which sequences are unanimously agreed to be real or fake?
- Which are the most contentious? Why?
- Were there any sequences that most students determined incorrectly? Why?

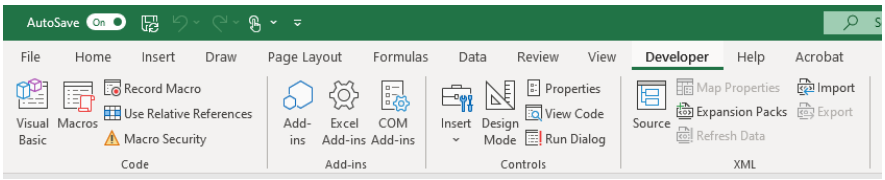
Making a Macro

Name: _____

To create a macro, first we need to make the developer menu visible. For Excel 2013 and 2010, select **File**, then **Options**. Choose **Customize Ribbon** on the left and then select the **Developer** checkbox on the right.



Now click OK. You should now be able to see the Developer tab.



Everything you do between clicking the 'Record Macro' and 'Stop Recording' button can be automatically repeated via a keyboard shortcut. For example, if you were using this spreadsheet:

	A	B	C	D	E	F	G	H
1	**Coinflip sequence**	**The same as the previous flip?**	**How long is the current streak?**			**Mean streak length**	**Maximum streak length**	**Total number of streaks**
2	H	Different	1			2.50	6	20
3	H	Same	2					
4	T	Different	1					
5	H	Different	1					
6	H	Same	2					
7	H	Same	3					

then a useful macro might be:

Step 1. Click 'Record Macro'. Input a shortcut key, such as Ctrl-m.
Step 2. Use F9/'Recalculate sheet' to generate a new set of coin flips.
Step 3. Copy the values in F2:H2 and paste them into F3:H3. Be sure to open the paste menu and select **paste as values** instead of pasting as a formula.
Step 4. Right-click **F3:H3** and select **Insert**... and **Shift cells down**.
Step 5. Click 'Stop Recording'.

Step 4 is important because the macro repeats your actions exactly. If you don't insert new cells, it will paste each new set of values on top of the old values.

If the macro is working, you should be able to hold down Ctrl-m to start automatically generating a string of results.

LESSON 2

re(Solve) MATHS BY INQUIRY Hot Streaks

Do Hands Get Hot?

Y9

About this lesson

Students learn about the hot hand phenomenon in basketball. Using their findings from Lesson 1: Pick the Fake Data, students explore different methodologies to discuss whether the phenomenon exists.

Australian Curriculum: Mathematics (Year 9)

ACMSP226: Calculate relative frequencies from given or collected data to estimate probabilities of events involving 'and' or 'or'.

Mathematical purpose

- Students work with a spreadsheet containing real basketball data and experiment with appropriate strategies to identify whether the hot hand effect is real.

Learning intention

- Is there such a thing as a 'hot hand' effect?

Time

Two lessons of approximately 1 hour each.

Vocabulary

- hot hand/hot streak
- random

Resources

- Student Sheet 1 – Garnett Sequence
- reSolve Excel Spreadsheet *2a Garnett Results Sample* (optional)
- reSolve Excel Spreadsheet *2b Garnett Results 2000–16*
- reSolve Excel Spreadsheet *2c Garnett Results Annotated*
- reSolve Excel Spreadsheet *2d Random Results*

reSolve: Mathematics by Inquiry is funded by the Australian Government and is managed by the Australian Academy of Science.

Teacher background information

The hot hand/hot streak in basketball refers to the idea that a player, after making a successful shot, becomes more likely to make another successful shot. As a result, when a player is on a 'streak', it becomes nearly impossible for them to miss a shot. The hot hand effect is widely believed to exist by players and fans, but statisticians have been debating its existence since the 1980s.

The appropriate methodology for identifying the hot hand is also debated (see Further reading in the Introduction). In this lesson we suggest a simplified approach, but teachers and students are encouraged to identify the flaws in this approach and to create their own more nuanced strategies.

Two NBA players are referred to in this resource: Klay Thompson and Kevin Garnett. A recording of Klay Thompson setting the NBA record for the most points scored in a quarter by scoring 13 consecutive throws is used to open the lesson and spark interest in the hot hand effect. The data investigated in the rest of the lesson come from Kevin Garnett's 16-year NBA career. Garnett's data were chosen for this purpose because his field goal average is 49.7%, making his field goal data comparable with the coin flipping data collected in Lesson 1.

Does the shot miss?

In this video ('Klay Thompson 37pt 3rd Quarter CSN Bay Area feed 1-23-15'), the NBA player Klay Thompson (#11) scores a record 37 points in a single quarter. Start playing the video and pause 15 seconds in immediately before Thompson takes his first shot. Ask students to stand if they believe the shot will go in or to stay seated if they believe it will miss. Continue playing and repeat for the next three shots (at 0:27, 0:44, 1:01).

Immediately after Thompson takes his fifth shot at 1:50, pause the video and explain the concept of the hot hand to the class (see Teacher background information). Ask students whether they think the hot hand effect exists and discuss different ways you could use basketball data to identify its existence.

Looking at basketball data

Resources: Student Sheet 1 – Garnett Sequence shows a sequence of 103 basketball goal attempts from 4 months of games by NBA All-Star player Kevin Garnett.

H represents scoring a goal (a hit) and M represents taking a shot and missing (a miss). This sequence is also available as a sequence of 103 Ms and Hs on the reSolve Excel Spreadsheet *2a Garnett Results Sample*. Inform students that, over his long career, Garnett's field goal average was 49.7%. Discuss with students why this is important if we are to compare results with random coin flips.

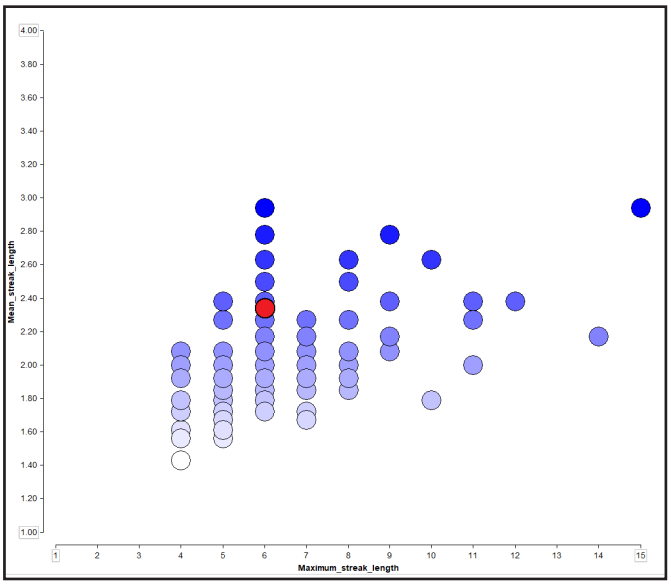
Resources: Ask students to analyse Garnett's maximum streak length and mean streak length using, for example, reSolve Excel Spreadsheet *1b Annotated Coin Flip Analyser* from Lesson 1.

Discuss with students how the maximum streak length and mean strength length would compare with random coin flips if the hot hand effect exists.

Ask: *Does the evidence suggest the existence of a hot streak in the data?*

Teacher notes:

- The hot hand effect suggests that multiple sequential hits in basketball are more likely than flipping sequential heads on a coin because each successful hit increases the chance of the next hit being successful.
- If the hot hand effect exists, we would expect the maximum and mean streak lengths to *not* look like they were randomly generated. They should look 'fake' when compared with the coin flip results.
- The mean streak length of the Garnett sequence is 2.34 and the maximum streak length is 6. The graph on the right shows the result of 100 randomly generated coin flip sequences (from Lesson 1) and highlights the datapoint of the Garnett sequence in red. From this it is plausible to conclude that the basketball results are *random*.



Ask students: *What might change if you were working with more data? What would you expect to stay the same?*

Searching for hot streaks

Resources: Have students open reSolve Excel Spreadsheet *2b Garnett Results 2000–16*.

Explain that each sheet in the file is a different NBA season and contains a sequential record of every shot Garnett took during that season. In total, there are 17 521 shots recorded.

Ask students to choose a single season to work on, individually or in pairs, ensuring that all seasons are covered. You may instead choose to designate students to seasons, noting that 2013–14, 2014–15 and 2015–16 contain significantly fewer data than the other seasons (as Garnett played in fewer games).

Each sheet in reSolve Excel Spreadsheet *2b Garnett Results 2000–16* is designed so that students can input strings of results (e.g. MHM) and it will output the number of times the string occurred in the season.

Prompt students to experiment with the spreadsheet by asking:

- *What is the longest streak of hits in the season?*
- *What is the longest streak of misses in the season?*
- *What is the longest streak of alternating hits and misses (i.e. HMHMHMHM) in the season?*

Note for students that the spreadsheet cannot identify overlapping strings; for example, when searching for MHM, the sequence MHMHM will be counted only once, but it will count twice if HM or MH is searched.

Discuss with the class the following strategies for finding streak lengths and the probabilities of making a successful shot.

Calculating streak lengths

- *A hot hand means that a player becomes more likely to make a successful shot if their previous shot was successful. How can we count how often a successful shot is followed by another successful shot?*
- *What string should we input to count the number of first hits in a streak?*
  - ◊ Searching MH will return the number of first hits in a streak; that is, the number of times a new streak begins. If the first shot of the season is a hit, it will not be counted, so 1 should be added to this value.
- *What string should we input to count the number of streaks in which the first two shots are hits?*
  - ◊ Searching MHH will return the number of streaks in which the first two shots are hits. Subtracting this from the number of MH streaks gives the number of streaks of hits of length exactly 1.
  - ◊ A similar process can provide the number of strings of length exactly 2, 3, 4, …

Calculating probabilities of successful shots, given previous success

- *If the first shot is a hit, how can we calculate the chance that the next shot is a hit (i.e. the chance the streak will continue)?*
  - ◊ If the next shot is a hit, the string will be MHH. Dividing the number of MHHs by the number of MHs will tell us the probability that an MH will be followed by another H.
    - ■ Students might suggest calculating the number of MHHs and comparing it with the number of MHMs. While this is a good strategy in principle, it should not be used here because the spreadsheet cannot identify overlapping strings (as discussed above).
- *If the second shot is also a hit, how can we calculate the chance that the third shot is a hit?*
  - ◊ We can find the probability that an MHH will be followed by another H by dividing the number of MHHHs by the number of MHHs.

Students should continue to apply this strategy until every string of hits has been accounted for.

Resources: See reSolve Excel Spreadsheet *2c Garnett Results Annotated* for a completed example (included for teacher reference).

Enabling prompts:

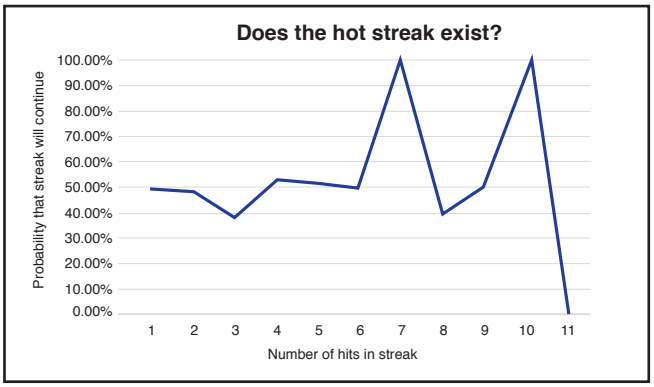
- *If the sequence were truly random, what would you expect the percentage chance to look like for each streak length?* [Approximately 50%.]
- *If the hot hand effect is real, what would you expect the percentage chance to look like for each streak length?* [You would expect the percentage chance to increase, signifying that the player becomes more likely to score if their previous shot scored.]
- *Why is there sometimes a 100% chance of a streak continuing? Does that mean it is impossible to miss?* [This is a result of the limited dataset each student uses, which results in very small sample sizes for long streak lengths. For example, if Garnett has streaks of 8 hits, but no streaks of exactly 7 hits, then the chance of a streak of 7 hits continuing will be given as 100%.]
- *If the hot hand effect exists, is it possible to eventually get so hot that you never miss again?*

Have students generate a graph that compares the probability of a streak continuing with the length of the streak. Discuss findings. How do results vary across seasons? Were there any seasons with outlying results?

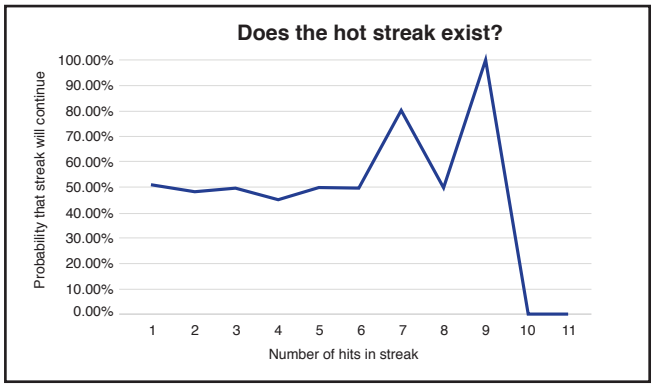
How random is basketball?

Resources: Explain that reSolve Excel Spreadsheet *2d Random Results* is identical to reSolve Excel Spreadsheet *2b Garnett Results 2000–16*, **except** that the basketball data have been replaced with computer-generated shots using Garnett's 49.7% average (i.e. every shot has a 49.7% chance of being a hit). Pressing F9 will generate new results.

Ask students to copy all their calculations across into the same season in reSolve Excel Spreadsheet *2c Garnett Results Annotated*, then compare the graph for these results with those of the Garnett graph. For example:



1 Graph from Garnett 2000–1 data



2 Graph from Random Results 2000–1 data

Discuss findings as a class:

- *Does the graph based on the random values look how you would expect?*
- *Every student will have a unique randomly generated graph. Are they similar or all distinctly different?*
- *How do they compare with the Garnett graph?*

Although the graphs all look different, there are some commonalities:

- Towards the end, as long as we have exceeded the maximum streak length, the probability of a successful shot will drop to 0.
- At the start, the probabilities are around 50% and are relatively consistent for streaks of length 1, 2, 3 or 4.
- Once the streaks have reached length 5 or more, the probabilities fluctuate a lot more because the sample size has decreased.

Reflection

Discuss as a class: *Do you think the hot hand effect exists? Why or why not?*

Finish watching the Thompson video, repeating the sit down/stand up exercise for each shot.

Further activity

The term 'heat check' is used to describe a tricky shot a basketball player attempts after making several easier shots. How might it have got this name? What effect might a heat check have on a hot streak?

Garnett Sequence

Name: _____

Below is Kevin Garnett's shot sequence from the start of the 2000–2001 NBA season.

H M H M M H M H H H M M
M H H M M M M M H H M M
H H M M M H M M M H H H
H H M M H M M H H H M M
H M H H M H H H H H H M M
M H H M M M M H M M H H
M M H H M M H H M M M M
H H M M H H H M M M M H
H H H M M M

Does this sequence appear to be randomly generated?