JOURNAL AND PROCEEDINGS
of
YOUNG ARCHIMEDESES

Volume 2, Number 1
2016

“Providing a forum to exchange mathematical ideas, activities, and/or sharing and interpreting high school research.”
## Contents

3. **Osman, F., Myadera, R.**
   - Editorial Board

7. **Scott, S.**
   - Pythagorean Triples

11. **Reece, J.H., Azleya, T.C.**
    - Primes of the Forms $n^2 - 2$

24. **Solano, M.L.**
    - Friends, Birthdays and Probability: Problems Concerning the Probability of Birthday Coincidences?

40. **Serb, A.**
    - What is the Relationship between the Speed of a falling object and the size and shape of the Crater caused by its impact with the ground?

50. **Innes, F.**
    - There cannot be an Odd Perfect Number using the Properties of Odds and Evens

---

The Editor, Journal and Proceedings of Young Archimedes
Trinity Grammar School,
119 Prospect Road,
SUMMER HILL NSW AUSTRALIA 2130

Published: January 2016
EDITORIAL BOARD

Dr Frederick Osman and Dr Ryohei Miyadera

BACKGROUND: On the 24th of September 2013, the Trinity Grammar School Music and Mathematics tour departed from Sydney International Airport and began a journey to Kwansei Gakuin Senior High School, Nishinomiya, Japan.

The main purpose of this tour was to establish a relationship with Kwansei Gakuin to develop an international programme that would go beyond the current Rugby connection, to promote cross-cultural awareness through extensive exchange programmes that challenge the mind, body and spirit for both students and staff from both schools.

JOURNAL NAME: Archimedes was a Greek mathematician, physicist, engineer, inventor, and astronomer. Archimedes is generally considered to be the greatest mathematician of antiquity and one of the greatest of all time.

AIMS:

1. The Journal and Proceedings of Young Archimedes publishes academic online papers of secondary students in the fields of Mathematics Applications.
2. To provide a forum to exchange mathematical ideas, activities, and/or sharing and interpreting high school research.
3. To pioneer a new field of educational endeavour to be the first Mathematics International Journal publication for High Schoolers.
4. To increase the relationship and strengthen the academic links between Trinity and Kwansei Gakuin.
5. To promote cross-cultural understanding between Australia and Japan and affirm our academic relationship as brother Schools.
6. To have students in all departments completing HSC and/or International Baccalaureate essays or projects with relevance to the fields of Mathematics Applications submit a paper for refereeing within the Journal and Proceedings of Young Archimedes.

OUTCOMES:

1. Issues are scheduled to be published in June and December of each Year.
2. Maximum of six long papers (max 6 pages) or twelve short papers (max 3 pages) for each issue.
3. An electronic online version of each issue is to be posted to the Trinity Grammar School Mathematics Club web site publication.
Journal and Proceedings of Young Archimedes

The Journal and Proceedings of Young Archimedes publishes academic online papers of secondary students in the fields of Mathematics Applications and provides a forum to exchange mathematical ideas, activities, and/or sharing and interpreting high school research.

Manuscripts will be reviewed by the Editor, in consultation with the Associate Editors, to decide whether the paper will be considered for publication in the Journal. Issues are scheduled to be published in June and December. An electronic version of each issue is posted to the Trinity Grammar School Mathematics Club web site http://bit.ly/tgs_archimedes as a formal publication. Enquiries relating to copyright or reproduction of an article should be directed to the author.

Information for Authors

Manuscripts are only accepted in digital format and should be e-mailed:
> In Australia to Dr Frederick Osman on fosman@trinity.nsw.edu.au or
> In Japan to Dr Ryohei Miyadera on runners@kwansei.ac.jp

The template below should be used as a guide for preparing manuscripts.
If the file-size is too large to email, it should be placed on a CD-ROM or other digital media and posted to:

The Editor, Journal of Proceedings of Young Archimedes
Trinity Grammar School,
119 Prospect Road,
SUMMER HILL NSW AUSTRALIA 2130

Editorial Board

Dr Frederick Osman has had an extensive experience of more than 20 years academic/industry experience in innovative teaching and researching, in Physics and Mathematics education. His research background and achievements have been attained in laser plasma interaction for inertial confinement fusion including work on several plasma effects. He is currently the Director of Vocational Education and the Master in Charge of the Mathematics Club at Trinity Grammar School.

Dr Ryohei Miyadera received a Ph.D. in Mathematics at Osaka City University and received a second Ph.D. in mathematics education at Kobe University. He has two fields of research: probability theory of functions with values in an abstract space and applications of Mathematica to discrete mathematics. He and his high school students have been doing research in discrete mathematics for more than 15 years.
ASSOCIATE EDITORS

Professor Robert Cowen is Mathematics Emeritus at Queens College, CUNY. He uses Mathematica in his own research and has written a textbook with John Kennedy called Discovering Mathematics with Mathematica.

Professor Heinrich Hora is known for his work on the theory for fusion energy with lasers. He has published more than 450 papers on laser-plasma interaction and inertial nuclear fusion, ponderomotive and relativistic self-focusing, laser acceleration of particles, correspondence principle of electromagnetic interaction and accuracy principle of nonlinearity.

Professor Yaichi Shinohara is Mathematics Emeritus of Kwansei Gakuin University. His research background and achievements have been in on Topology and Knot theory. He is a respected mathematician in the fields of algebra and geometry.

Professor Tadashi Takahashi is a Mathematics Professor at Konan University. He is the President of Japan Society for Symbolic and Algebraic Computation and is also the President of the Game Amusement Society. His research background and achievements have been in Computer Algebra, Singularity Theory and Efficient Use of Computer in Mathematics Education.

Dr Jonny Bernas Pornel is an Assistant Professor of Mathematics in the University of the Philippines Visayas. He promotes the use of Lesson Study among teachers of Science and Mathematics and the enhancement of mathematical creativity among students.

Dr Nethal K. Jajo is Modelling and Projection Analyst at the University of Sydney, Australia. He has a PhD degree in Mathematical Statistics and Probability Theory with professional training in Discrete Event Simulation and System Dynamics Modelling. His industrial and academic research activities include: Data analysis, data mining, partial least squares path modelling, regression analysis and dynamic simulation.

Dr Katsuyuki Yoshikawa is an International expert in the area of Knot Theory. He has attained the Takebe-Award for his research on the four dimensional topology from the Mathematical Society of Japan.

Edward Habkouk is an experienced teacher of NSW Mathematics courses to HSC level and he is an HSC Mathematics Extension 1 marker (since 2000) and IB Mathematics Examiner (particularly SL, paper 1) since 1999. He is currently the Dean of Mathematics at Trinity Grammar School.

Stephen McAndrew is a former Physics Teacher at Trinity Grammar School, having taught in Australia and the UK. His research background is in Applied Mathematics, in particular the areas of classical mechanics, fluid mechanics and electromagnetism. He is currently involved in a PhD research in magnetohydrodynamic shock waves.

Katsuya Mori is a teacher at Takarazuka Higashi High School who is doing research in Mathematics with his students. His students papers were published at The Rose-Hulman Undergraduate Mathematics Journal. He achieved a M.Sc. from Kyoto University with a major in algebraic geometry.

Shane Scott is an experienced teacher of NSW Mathematics courses to HSC level at Trinity Grammar School. He is an executive member of the Mathematical Association of New South Wales. He has won the NSW Premier’s Scholarship for Mathematical Teaching and has travelled to Germany, UK and the US to attend and present at International Schools and Mathematics conferences.

Yuko Matsuda is known as an experienced computer scientist with many years’ experience in artificial intelligence, data science, language design and super-computing based on symbolic computation.
Sample template format for Authors to the Journal and Proceedings of Young Archimedes (16 point font size)

First name Last name

Department, Institution, City, Country

E-mail: name@domain-name

Abstract

This is the layout and template for a paper to be submitted to The Journal and Proceedings of Young Archimedes.

Introduction (12 point font size)

The Journal and Proceedings of Young Archimedes publishes academic online papers of secondary students in the fields of Mathematics Applications and provides a forum to exchange of mathematical ideas, activities, and or sharing and interpreting high school research. Papers may be submitted electronically only to the editors Dr Frederick Osman on fosman@trinity.nsw.edu.au from Trinity Grammar School Australia and Dr Ryohei Miyadera on runners@kwansei.ac.jp from Kwansei Gakuin High School, Nishinomiya City Japan.

Acknowledgement of receipt of the submission will be sent to the corresponding author’s e-mail address. It is the author’s responsibility to submit an accurate manuscript – any errors in spelling, grammar, or scientific content may be reproduced as typed by the author. Manuscripts will be reviewed by the Editor, in consultation with the Associate Editors, to decide whether the paper will be considered for publication in the Journal. Accepted papers will be published electronically on the Trinity Grammar School Mathematics Club web-site.

Layout and style

A Times New Roman font is used for the main text. The font size is 11 points with main heading of sections should use font size 12 points. It is important that when the final PDF file is created, all fonts used must be embedded. Two columns are used except for the title and abstract section and possibly for large figures, tables or photographs that need a full-page width. If you have any questions regarding paper submission, please contact the editors, Dr Frederick Osman from Trinity Grammar School, Sydney AUSTRALIA and Dr Ryohei Miyadera from Kwansei Gakuin High School, Nishinomiya City JAPAN.

Equations (11 point font size)

Equations should be placed on separate lines and numbered. An example of an equation is given below:

\[ f = -\nabla p + f_N, \]  

where \( f_N \) is the nonlinear force (Osman, 2000).

Figures

All figures must be centered on the column (or page, if the figure spans both columns).

Figure 1: Generation of blocks of deuterium plasma moving against the neodymium glass laser light (Osman, 2005).

References (11 point font size)


Pythagorean Triples

Shane Scott

Trinity Grammar School, Mathematics Department, Summer Hill NSW, Australia

E-mail: sscott@trinity.nsw.edu.au

Abstract

From early Secondary School we are introduced to Pythagoras’ Theorem – The square of the hypotenuse of a right triangle is equal to the sum of the squares of the two adjacent sides, and the associated formula \( c^2 = a^2 + b^2 \). In school textbooks it has been used to find the length of a side of a triangle. A little part of the textbook is put aside to mention the converse of the theorem; that is if you have a triangle and the sum of the squares of two adjacent sides is equal to the square of the longest side then you must have a right angled triangle. The application of the converse has been used throughout history, from the Babylonians in which the tablet Plimpton 322 shows a collection of Pythagorean Triples. The Egyptians used them to ensure the blocks for the pyramids were perfectly square.

How many Pythagorean Triples are there? Is there finitely many or infinitely many? If there are infinitely many are there infinitely many different ratios (or primitives) of triples?

We know \( \{3, 4, 5\} \) is a triple and because we can multiply these by any natural number and generate a new triple. So we now know that we have infinitely many possibilities.

Using a fibonacci numbers to generate Pythagorean Triples?

To generate a Pythagorean triple \( \{x, y, z\} \) take any sequence of 4 consecutive numbers \( a, b, c, d \) and do the following:-

\[
x = a \times d
\]

\[
y = 2 \times b \times c
\]

\[
z = b^2 + c^2
\]

Given any two starting numbers \( a \) and \( b \), if a recurring sequence is generated a Pythagorean Triple can be generated using the above rules. Can we prove it? Yes we can. Let’s look at the first four terms of the sequence (this would be equivalent to taking any two consecutive terms in a recurring sequence). They are \( a, b, a+b, a+2b \)

Now generate \( x, y \) and \( z \)

\[
x = a \times (a + 2b)
\]

\[
z = a^2 + 2ab
\]

\[
y = 2 \times b \times (a + b)
\]

\[
y = 2ab + 2b^2
\]

\[
z = b^2 + (a + b)^2
\]

We now need to show

\[
x^2 + y^2 = z^2 \text{ i.e.} \]

\[
(a^2 + 2ab)^2 + (2ab + 2b^2)^2 = (b^2 + (a + b)^2)^2
\]

L.H.S

\[
(a^2 + 2ab)^2 + (2ab + 2b^2)^2
\]

\[
= a^4 + 4a^3b + 4a^2b^2 + 8ab^3 + 4b^4
\]

\[
= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4 + 2a^2b^2 + 4ab^3 + 3b^4
\]

\[
= (a + b)^4 + b^4 + 2b^2(a + b)^2
\]

\[
= (b^2 + (a + b)^2)^2
\]

= R.H.S
Therefore any set of 4 consecutive numbers of a recurring sequence will generate Pythagorean Triples.

**Another more traditional way of generating triples**

Let us consider a set of a Pythagorean Triple \( \{a, b, c\} \), we know that \( a, b \) and \( c \) cannot all be odd.

**Proof by contradiction:**

Let’s assume that \( a, b \) and \( c \) are odd. Then \( a^2, b^2 \) and \( c^2 \) are also odd.

But an odd number plus an odd number must be an even number. Therefore \( c^2 \) cannot be odd and hence \( c \) cannot be odd.

An equivalent statement to the one above is that at least one of \( a, b \) or \( c \) must be even.

So we can write \( a = 2mn \) where \( m \) and \( n \) are natural numbers. Using Pythagoras’ theorem we have:

\[
\begin{align*}
2c &= 2m^2 + 2n^2 \\
c &= m^2 + n^2 
\end{align*}
\]

Subtracting the two equations gives:

\[
\begin{align*}
2b &= 2m^2 - 2n^2 \\
b &= m^2 - n^2 
\end{align*}
\]

So by setting:

\[
\begin{align*}
a &= 2mn \\
b &= m^2 + n^2 
\end{align*}
\]

And \( c = m^2 + n^2 \)

We have a method of generating infinitely many Pythagorean Triples.

**A Geometrical Look:**

Let’s start with the formula

\[
a^2 + b^2 = c^2
\]

If we divide by \( c^2 \) we get:

\[
1 = \left( \frac{a}{c} \right)^2 + \left( \frac{b}{c} \right)^2
\]

If \( a, b \) and \( c \) are natural numbers then:

\[
\left( \frac{a}{c} \right)^2 \text{ and } \left( \frac{b}{c} \right)^2 \text{ must be rational numbers.}
\]

However if \( 1 = \left( \frac{a}{c} \right)^2 + \left( \frac{b}{c} \right)^2 \) then we can say that \( \left( \frac{a}{c}, \frac{b}{c} \right) \) lie on the unit circle.
This means there is a direct relationship between rational points on the unit circle and Pythagorean triples and hence all primitive Pythagorean triples are represented as rational points on the unit circle.

To find these points we need to look at the intersection of the line $y = M(x+1)$. The capital $M$ is used for the gradient to avoid confusion with the $m$ used in the formula to generate triples and the unit circle $x^2 + y^2 = 1$

Let us firstly look at the line $y = M(x+1)$. As $M$ changes the gradient of the line changes, however all lines of this form pass through the point $(-1,0)$ and another rational point in the first quadrant.

To find the second point, we need to solve simultaneous equations $x^2 + y^2 = 1$ and $y = M(x+1)$

Substituting $y = M(x+1)$ into $x^2 + y^2 = 1$ we get $x^2 + [M(x+1)]^2 = 1$

When we expand it becomes $(1 + M^2)x^2 + (2M^2)x + (M^2 - 1) = 0$

This can be factorised to be $(x+1)((1 + M^2)x + (M^2 - 1)) = 0$

This gives us two solutions

If $x+1 = 0$, then $x = -1$ and $y = 0$ which gives us the point $(-1, 0)$

Using the solution $(1 + M^2)x + (M^2 - 1) = 0$ we get $x = \frac{1 - M^2}{1 + M^2}$ and the corresponding $y = \frac{2M}{1 + M^2}$

For the line $y = M(x+1)$ we will now substitute $M$ for $\frac{n}{m}$ where $m > n$ and $m, n$ are natural numbers.

This allows us to write the second point of intersection $\left(\frac{1 - M^2}{1 + M^2}, \frac{2M}{1 + M^2}\right)$ as
\[
\frac{1 - \left(\frac{n}{m}\right)^2}{1 + \left(\frac{n}{m}\right)^2} \cdot \frac{2\left(\frac{n}{m}\right)}{1 + \left(\frac{n}{m}\right)^2}
\]
which simplifies to
\[
\frac{\left(m^2 - n^2\right)}{m^2 + n^2} \cdot \frac{2mn}{m^2 + n^2}
\]
At this point is on the unit circle it implies
\[
1 = \left(\frac{m^2 - n^2}{m^2 + n^2}\right) + \left(\frac{2mn}{m^2 + n^2}\right)^2
\]
And so we can now say
\[
a = 2mn \\
b = m^2 - n^2
\]
and
\[
c = m^2 + n^2
\]
This means that not only do we have a method of constructing infinitely many Pythagorean Triples it is a method of constructing all the primitives of Pythagorean Triples.
Primes of the Form \( n^2 - 2 \)

Jeb Reece H. Grabato, Azleya Trish C. Canuto, Pearly Jane A. Cortez
Keneth Noe H. Escueta, Agham Bayan B. Martinete, Gercel Therese R. Serafino
Advisers: Prof. Jonny Pornel, Prof. Raphael Belleza,
Prof. Early Sol Gadong, Prof. Mary Anne Naragdao
UP High School in Iloilo, University of the Philippines Visayas, Iloilo City Philippines

This paper presented the concept of primes with the form \( P = n^2 - 2 \) and determined its properties. It found the following characteristics of \( P \): (a) \( P \) is congruent to 3, 7 or 9 modulo 10; (b) any \( P \neq 7 \) will never have a twin prime; (c) \( P > 2 \) are congruent to 3 modulo 4; and (d) \( P > 2 \) are congruent to 7 modulo 8. It also found the following characteristics of \( n \): (a) \( n > 2 \) are odd; (b) \( n \neq 3 \) is not congruent to \( \pm 3 \) modulo 7; and (c) \( n \) is not congruent to \( \pm 6 \) modulo 17. Lastly, it found that the prime factor \( p \) of a non-prime \( n^2 - 2 \) is congruent to \( \pm 1 \) modulo 8.

Key words: primes of the form \( P = n^2 - 2 \)

Introduction

In 1978, Shanks conjectured in his book that there are infinitely many primes with the form \( n^2 - 2 \). “While more than 15,000 of such primes are known... a proof of the conjecture is still awaited” (Shanks, 1978).

This paper aims to provide information about the characteristics of \( n \) that will make \( n^2 - 2 \) prime, as well as the characteristics of a prime number with the form \( n^2 - 2 \).

Problem

This mathematical investigation aims to answer the following questions:

1) What are the characteristics of \( n \) that will make \( P = n^2 - 2 \) prime?
2) What are the characteristics of prime numbers of the form \( P = n^2 - 2 \)?
3) What are the properties of the prime factors of non-prime \( n^2 - 2 \)?

Conjectures Formulated

To come up with a conjecture, the researchers identified the first few primes with the form \( n^2 - 2 \). Next, their characteristics were observed and determined.
Table 1: Values of $n < 65$ that make $n^2 - 2$ prime.

<table>
<thead>
<tr>
<th>$n$</th>
<th>Prime $P$</th>
<th>$n(mod 7)$</th>
<th>$n(mod 17)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>-2</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>47</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>79</td>
<td>2</td>
<td>-8</td>
</tr>
<tr>
<td>13</td>
<td>167</td>
<td>-1</td>
<td>-4</td>
</tr>
<tr>
<td>15</td>
<td>223</td>
<td>1</td>
<td>-2</td>
</tr>
<tr>
<td>19</td>
<td>359</td>
<td>-2</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>439</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>27</td>
<td>727</td>
<td>-1</td>
<td>-7</td>
</tr>
<tr>
<td>29</td>
<td>839</td>
<td>1</td>
<td>-5</td>
</tr>
<tr>
<td>33</td>
<td>1087</td>
<td>-2</td>
<td>-1</td>
</tr>
<tr>
<td>35</td>
<td>1223</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>37</td>
<td>1367</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>43</td>
<td>1847</td>
<td>1</td>
<td>-8</td>
</tr>
<tr>
<td>47</td>
<td>2207</td>
<td>-2</td>
<td>-4</td>
</tr>
<tr>
<td>49</td>
<td>2399</td>
<td>0</td>
<td>-2</td>
</tr>
<tr>
<td>55</td>
<td>3023</td>
<td>-1</td>
<td>4</td>
</tr>
<tr>
<td>61</td>
<td>3719</td>
<td>-2</td>
<td>-7</td>
</tr>
<tr>
<td>63</td>
<td>3967</td>
<td>0</td>
<td>-5</td>
</tr>
</tbody>
</table>

The values of the first 20 $n$ that make $n^2 - 2$ prime are shown in Table 1. All $n$ greater than 2 are odd. Thus, Conjecture 1 was formulated.

**Conjecture 1:** If $n>2$ and $P = n^2 - 2$ is a prime, then $n$ is odd.

It can also be seen in Table 1 that the possible remainders of $n \div 7$, with the exception of $n = 3$, can only be 0, ±1, and ±2. Also, the table shows that $n$ cannot be congruent to $\pm 6(mod 17)$. The table may not show that $n$ can be 0(mod17), but this will be confirmed later in Table 7. Thus, Conjectures 2 and 3 were formulated.

**Conjecture 2:** If $P = n^2 - 2$ is prime and $n \neq 3$, then $n \neq \pm 3(mod 7)$

**Conjecture 3:** If $P = n^2 - 2$ is prime, then $n \neq \pm 6(mod 17)$.

Table 2: Table of Values for $P \pm 2$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Prime $P = n^2 - 2$</th>
<th>$P + 2$</th>
<th>$P - 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>47</td>
<td>49</td>
<td>45</td>
</tr>
<tr>
<td>9</td>
<td>79</td>
<td>81</td>
<td>77</td>
</tr>
<tr>
<td>13</td>
<td>167</td>
<td>169</td>
<td>165</td>
</tr>
<tr>
<td>15</td>
<td>223</td>
<td>225</td>
<td>221</td>
</tr>
<tr>
<td>19</td>
<td>359</td>
<td>361</td>
<td>357</td>
</tr>
<tr>
<td>21</td>
<td>439</td>
<td>441</td>
<td>437</td>
</tr>
<tr>
<td>27</td>
<td>727</td>
<td>729</td>
<td>725</td>
</tr>
<tr>
<td>29</td>
<td>839</td>
<td>841</td>
<td>837</td>
</tr>
<tr>
<td>33</td>
<td>1087</td>
<td>1089</td>
<td>1085</td>
</tr>
<tr>
<td>35</td>
<td>1223</td>
<td>1225</td>
<td>1221</td>
</tr>
<tr>
<td>37</td>
<td>1367</td>
<td>1369</td>
<td>1365</td>
</tr>
<tr>
<td>43</td>
<td>1847</td>
<td>1849</td>
<td>1845</td>
</tr>
<tr>
<td>47</td>
<td>2207</td>
<td>2209</td>
<td>2205</td>
</tr>
<tr>
<td>49</td>
<td>2399</td>
<td>2401</td>
<td>2397</td>
</tr>
<tr>
<td>55</td>
<td>3023</td>
<td>3025</td>
<td>3021</td>
</tr>
<tr>
<td>61</td>
<td>3719</td>
<td>3721</td>
<td>3717</td>
</tr>
<tr>
<td>63</td>
<td>3967</td>
<td>3969</td>
<td>3965</td>
</tr>
</tbody>
</table>

Table 2 shows that all primes of the form $n^2 - 2$ that are greater than 2 can only end with 3, 7, or 9. Thus Conjecture 4 logically follows.

**Conjecture 4:** Let $P = n^2 - 2$ be prime. If $P>2$, then $P \equiv d(mod 10)$ where $d\in\{3, 7, 9\}$.

Also, it can be seen in Table 2 that the values for $P - 2$ and $P + 2$ are not primes, with the exception of $P = 7$ because $7 - 2 = 5$, which is a prime. Thus, Conjecture 5 was advanced.
Conjecture 5: If \( P = n^2 - 2 \) is prime and \( P \neq 7 \), then \( P \) has no twin prime.

Table 3: \( P \mod m \), where \( 0 < P < 3970 \) and \( m \in \{4, 8\} \)

<table>
<thead>
<tr>
<th>Prime ( P = n^2 - 2 )</th>
<th>( P \mod 4 )</th>
<th>( P \mod 8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>47</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>79</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>167</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>223</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>359</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>439</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>727</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>839</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>1087</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>1223</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>1367</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>1847</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>2207</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>2399</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>3023</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>3719</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>3967</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

In Table 3, it can be seen that all \( P > 2 \) are congruent to \( 3 \mod 4 \) and to \( 7 \mod 8 \). Thus, Conjectures 6 and 7 logically follow.

Conjecture 6: If \( P = n^2 - 2 \) is prime and \( P > 2 \), then \( P \equiv 3 \mod 4 \)

Conjecture 7: If \( P = n^2 - 2 \) is prime and \( P > 2 \), then \( P \equiv 7 \mod 8 \)

Table 4: Composite \( n^2 - 2 \) and its prime factor \( p \), where \( 2 < n < 80 \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( n^2 - 2 )</th>
<th>( p )</th>
<th>( p \mod 8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>119</td>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>17</td>
<td>287</td>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>23</td>
<td>527</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>623</td>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>31</td>
<td>959</td>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>39</td>
<td>1519</td>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>41</td>
<td>1679</td>
<td>23</td>
<td>-1</td>
</tr>
<tr>
<td>45</td>
<td>2023</td>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>51</td>
<td>2599</td>
<td>23</td>
<td>-1</td>
</tr>
<tr>
<td>53</td>
<td>2807</td>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>57</td>
<td>3247</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>59</td>
<td>3479</td>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>65</td>
<td>4223</td>
<td>41</td>
<td>1</td>
</tr>
<tr>
<td>67</td>
<td>4487</td>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>73</td>
<td>5327</td>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>79</td>
<td>6239</td>
<td>17</td>
<td>1</td>
</tr>
</tbody>
</table>

In Table 4, it can be seen that all non-prime \( n^2 - 2 \) have prime factors \( p \) that are congruent to \( \pm 1 \mod 8 \).

Thus, Conjecture 8 was formulated.

Conjecture 8: If \( n^2 - 2 \) is not prime, then its prime factor \( p \equiv \pm 1 \mod 8 \)
Verifying Conjectures

Table 5: Values of $800 < n < 1000$ that make $P = n^2 - 2$ prime

<table>
<thead>
<tr>
<th>Prime $P = n^2 - 2$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>651 247</td>
<td>807</td>
</tr>
<tr>
<td>657 719</td>
<td>811</td>
</tr>
<tr>
<td>667 487</td>
<td>817</td>
</tr>
<tr>
<td>680 623</td>
<td>825</td>
</tr>
<tr>
<td>707 279</td>
<td>841</td>
</tr>
<tr>
<td>744 767</td>
<td>863</td>
</tr>
<tr>
<td>765 623</td>
<td>875</td>
</tr>
<tr>
<td>776 159</td>
<td>881</td>
</tr>
<tr>
<td>804 607</td>
<td>897</td>
</tr>
<tr>
<td>811 799</td>
<td>901</td>
</tr>
<tr>
<td>866 759</td>
<td>931</td>
</tr>
<tr>
<td>889 247</td>
<td>943</td>
</tr>
<tr>
<td>893 023</td>
<td>945</td>
</tr>
<tr>
<td>904 399</td>
<td>951</td>
</tr>
<tr>
<td>919 679</td>
<td>959</td>
</tr>
<tr>
<td>946 727</td>
<td>973</td>
</tr>
<tr>
<td>958 439</td>
<td>979</td>
</tr>
<tr>
<td>974 167</td>
<td>987</td>
</tr>
<tr>
<td>986 047</td>
<td>993</td>
</tr>
<tr>
<td>990 023</td>
<td>995</td>
</tr>
</tbody>
</table>

The list of primes $P$ with the form $n^2 - 2$, where $800 < n < 1000$, is shown in Table 5. All values of $n$ are odd, thus making Conjecture 1 logical.

Table 6: $n(\text{mod 7})$, where $800 < n < 1000$

<table>
<thead>
<tr>
<th>Prime $P = n^2 - 2$</th>
<th>$n$</th>
<th>$n(\text{mod 7})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>651 247</td>
<td>807</td>
<td>2</td>
</tr>
<tr>
<td>657 719</td>
<td>811</td>
<td>-1</td>
</tr>
<tr>
<td>667 487</td>
<td>817</td>
<td>-2</td>
</tr>
<tr>
<td>680 623</td>
<td>825</td>
<td>-1</td>
</tr>
<tr>
<td>707 279</td>
<td>841</td>
<td>1</td>
</tr>
<tr>
<td>744 767</td>
<td>863</td>
<td>2</td>
</tr>
<tr>
<td>765 623</td>
<td>875</td>
<td>0</td>
</tr>
<tr>
<td>776 159</td>
<td>881</td>
<td>-1</td>
</tr>
<tr>
<td>804 607</td>
<td>897</td>
<td>1</td>
</tr>
<tr>
<td>811 799</td>
<td>901</td>
<td>-2</td>
</tr>
<tr>
<td>866 759</td>
<td>931</td>
<td>0</td>
</tr>
<tr>
<td>889 247</td>
<td>943</td>
<td>-2</td>
</tr>
<tr>
<td>893 023</td>
<td>945</td>
<td>0</td>
</tr>
<tr>
<td>904 399</td>
<td>951</td>
<td>-1</td>
</tr>
<tr>
<td>919 679</td>
<td>959</td>
<td>0</td>
</tr>
<tr>
<td>946 727</td>
<td>973</td>
<td>0</td>
</tr>
<tr>
<td>958 439</td>
<td>979</td>
<td>-1</td>
</tr>
<tr>
<td>974 167</td>
<td>987</td>
<td>0</td>
</tr>
<tr>
<td>986 047</td>
<td>993</td>
<td>-1</td>
</tr>
<tr>
<td>990 023</td>
<td>995</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6 shows that $800 < n < 1000$ cannot be $\pm 3(\text{mod 7})$

This makes Conjecture 2 logical.
Table 7: \( n(\text{mod 17}) \), where 
\[ 800 < n < 1000 \]

<table>
<thead>
<tr>
<th>Prime ( P = n^2 - 2 )</th>
<th>( n )</th>
<th>( n(\text{mod 17}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>651 247</td>
<td>807</td>
<td>8</td>
</tr>
<tr>
<td>657 719</td>
<td>811</td>
<td>-5</td>
</tr>
<tr>
<td>667 487</td>
<td>817</td>
<td>1</td>
</tr>
<tr>
<td>680 623</td>
<td>825</td>
<td>-8</td>
</tr>
<tr>
<td>707 279</td>
<td>841</td>
<td>8</td>
</tr>
<tr>
<td>744 767</td>
<td>863</td>
<td>-4</td>
</tr>
<tr>
<td>765 623</td>
<td>875</td>
<td>8</td>
</tr>
<tr>
<td>776 159</td>
<td>881</td>
<td>-3</td>
</tr>
<tr>
<td>804 607</td>
<td>897</td>
<td>-4</td>
</tr>
<tr>
<td>811 799</td>
<td>901</td>
<td>0</td>
</tr>
<tr>
<td>866 759</td>
<td>931</td>
<td>-4</td>
</tr>
<tr>
<td>889 247</td>
<td>943</td>
<td>8</td>
</tr>
<tr>
<td>893 023</td>
<td>945</td>
<td>-7</td>
</tr>
<tr>
<td>904 399</td>
<td>951</td>
<td>-1</td>
</tr>
<tr>
<td>919 679</td>
<td>959</td>
<td>7</td>
</tr>
<tr>
<td>946 727</td>
<td>973</td>
<td>4</td>
</tr>
<tr>
<td>958 439</td>
<td>979</td>
<td>-7</td>
</tr>
<tr>
<td>974 167</td>
<td>987</td>
<td>1</td>
</tr>
<tr>
<td>986 047</td>
<td>993</td>
<td>7</td>
</tr>
<tr>
<td>990 023</td>
<td>995</td>
<td>-8</td>
</tr>
</tbody>
</table>

Table 7 shows that \( 800 < n < 1000 \) cannot be \( \pm 6(\text{mod 17}) \). This makes Conjecture 3 logical.

Table 8: Prime \( P = n^2 - 2 \) where 
\[ 650 00 < P < 1 000 000 \]

<table>
<thead>
<tr>
<th>( n )</th>
<th>Prime ( P = n^2 - 2 )</th>
<th>Last digit of ( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>807</td>
<td>651 247</td>
<td>7</td>
</tr>
<tr>
<td>811</td>
<td>657 719</td>
<td>9</td>
</tr>
<tr>
<td>817</td>
<td>667 487</td>
<td>7</td>
</tr>
<tr>
<td>825</td>
<td>680 623</td>
<td>3</td>
</tr>
<tr>
<td>841</td>
<td>707 279</td>
<td>9</td>
</tr>
<tr>
<td>863</td>
<td>744 767</td>
<td>7</td>
</tr>
<tr>
<td>875</td>
<td>765 623</td>
<td>3</td>
</tr>
<tr>
<td>881</td>
<td>776 159</td>
<td>9</td>
</tr>
<tr>
<td>897</td>
<td>804 607</td>
<td>7</td>
</tr>
<tr>
<td>901</td>
<td>811 799</td>
<td>9</td>
</tr>
<tr>
<td>931</td>
<td>866 759</td>
<td>9</td>
</tr>
<tr>
<td>943</td>
<td>889 247</td>
<td>7</td>
</tr>
<tr>
<td>945</td>
<td>893 023</td>
<td>3</td>
</tr>
<tr>
<td>951</td>
<td>904 399</td>
<td>9</td>
</tr>
<tr>
<td>959</td>
<td>919 679</td>
<td>9</td>
</tr>
<tr>
<td>973</td>
<td>946 727</td>
<td>7</td>
</tr>
<tr>
<td>979</td>
<td>958 439</td>
<td>9</td>
</tr>
<tr>
<td>987</td>
<td>974 167</td>
<td>7</td>
</tr>
<tr>
<td>993</td>
<td>986 047</td>
<td>7</td>
</tr>
<tr>
<td>995</td>
<td>990 023</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 8 shows that all prime \( P = n^2 - 2 \) will always end with 3, 7, or 9. This verifies Conjecture 4.
Table 9: Table of Values for \( P \pm 2 \) where \( 650 \, 000 < P < 1 \, 000 \, 000 \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \text{Prime} ) ( P = n^2 - 2 )</th>
<th>( P + 2 )</th>
<th>( P - 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>807</td>
<td>651 247</td>
<td>651 249</td>
<td>651 245</td>
</tr>
<tr>
<td>811</td>
<td>657 719</td>
<td>657 721</td>
<td>657 717</td>
</tr>
<tr>
<td>817</td>
<td>667 487</td>
<td>667 489</td>
<td>667 485</td>
</tr>
<tr>
<td>825</td>
<td>680 623</td>
<td>680 625</td>
<td>680 621</td>
</tr>
<tr>
<td>841</td>
<td>707 279</td>
<td>707 281</td>
<td>707 277</td>
</tr>
<tr>
<td>863</td>
<td>744 767</td>
<td>744 769</td>
<td>744 765</td>
</tr>
<tr>
<td>875</td>
<td>765 623</td>
<td>765 625</td>
<td>765 621</td>
</tr>
<tr>
<td>881</td>
<td>776 159</td>
<td>776 161</td>
<td>776 157</td>
</tr>
<tr>
<td>897</td>
<td>804 607</td>
<td>804 609</td>
<td>804 605</td>
</tr>
<tr>
<td>901</td>
<td>811 799</td>
<td>811 801</td>
<td>811 797</td>
</tr>
<tr>
<td>931</td>
<td>866 759</td>
<td>866 761</td>
<td>866 757</td>
</tr>
<tr>
<td>943</td>
<td>889 247</td>
<td>889 249</td>
<td>889 245</td>
</tr>
<tr>
<td>945</td>
<td>893 023</td>
<td>893 025</td>
<td>893 021</td>
</tr>
<tr>
<td>951</td>
<td>904 399</td>
<td>904 401</td>
<td>904 397</td>
</tr>
<tr>
<td>959</td>
<td>919 679</td>
<td>919 681</td>
<td>919 677</td>
</tr>
<tr>
<td>973</td>
<td>946 727</td>
<td>946 729</td>
<td>946 725</td>
</tr>
<tr>
<td>979</td>
<td>958 439</td>
<td>958 441</td>
<td>958 437</td>
</tr>
<tr>
<td>987</td>
<td>974 167</td>
<td>974 169</td>
<td>974 165</td>
</tr>
<tr>
<td>993</td>
<td>986 047</td>
<td>986 049</td>
<td>986 045</td>
</tr>
<tr>
<td>995</td>
<td>990 023</td>
<td>990 025</td>
<td>990 021</td>
</tr>
</tbody>
</table>

Table 9 shows that all \( P + 2 \) and \( P - 2 \) are not primes. This verifies Conjecture 5.

Table 10: \( P \pmod{4} \), where \( 650 \, 000 < P < 1 \, 000 \, 000 \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \text{Prime} ) ( P = n^2 - 2 )</th>
<th>( P \pmod{4} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>807</td>
<td>651 247</td>
<td>3</td>
</tr>
<tr>
<td>811</td>
<td>657 719</td>
<td>3</td>
</tr>
<tr>
<td>817</td>
<td>667 487</td>
<td>3</td>
</tr>
<tr>
<td>825</td>
<td>680 623</td>
<td>3</td>
</tr>
<tr>
<td>841</td>
<td>707 279</td>
<td>3</td>
</tr>
<tr>
<td>863</td>
<td>744 767</td>
<td>3</td>
</tr>
<tr>
<td>875</td>
<td>765 623</td>
<td>3</td>
</tr>
<tr>
<td>881</td>
<td>776 159</td>
<td>3</td>
</tr>
<tr>
<td>897</td>
<td>804 607</td>
<td>3</td>
</tr>
<tr>
<td>901</td>
<td>811 799</td>
<td>3</td>
</tr>
<tr>
<td>931</td>
<td>866 759</td>
<td>3</td>
</tr>
<tr>
<td>943</td>
<td>889 247</td>
<td>3</td>
</tr>
<tr>
<td>945</td>
<td>893 023</td>
<td>3</td>
</tr>
<tr>
<td>951</td>
<td>904 399</td>
<td>3</td>
</tr>
<tr>
<td>959</td>
<td>919 679</td>
<td>3</td>
</tr>
<tr>
<td>973</td>
<td>946 727</td>
<td>3</td>
</tr>
<tr>
<td>979</td>
<td>958 439</td>
<td>3</td>
</tr>
<tr>
<td>987</td>
<td>974 167</td>
<td>3</td>
</tr>
<tr>
<td>993</td>
<td>986 047</td>
<td>3</td>
</tr>
<tr>
<td>995</td>
<td>990 023</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 10 shows that \( 650 \, 000 < P < 1 \, 000 \, 000 \) all congruent to \( 3 \pmod{4} \). This confirms Conjecture 6.
In Table 12, it can be seen that all non-prime $n^2 - 2$ have prime factors $p$ that are congruent to $\pm 1 \pmod{8}$. Thus, Conjecture 8 is verified.

To test the conjectures in extreme cases, let $n = 179965$. According to Shanks (1978), having 179 965 as the value of $n$ will make $P = n^2 - 2$ prime. Since $n$ is odd, this verifies Conjecture 1.

To verify Conjecture 2, it must be shown that 179 965 is not congruent to $\pm 3 \pmod{7}$. Since 179 963 is divisible by 7, this implies

\[179963 \equiv 0 \pmod{7}\]

\[179963 + 2 \equiv 0 \pmod{7} + 2\]

\[179965 \equiv 2 \pmod{7}\]

Since 179 965 $\equiv 2 \pmod{7}$, Conjecture 2 is verified.

Using the same argument to confirm Conjecture 2, it can be shown that 179 965 is not congruent to $\pm 6 \pmod{17}$. Since 179 962 is divisible by 17, this implies

\[179962 \equiv 0 \pmod{17}\]

\[179963 + 3 \equiv 0 \pmod{17} + 3\]

\[179965 \equiv 3 \pmod{17}\]
Since \(179,965 \equiv 3(\text{mod}17)\), Conjecture 3 is verified.

To verify Conjecture 4, it must be shown that \(P\) ends with 3, 7, or 9.

\[
P = n^2 - 2
\]

\[
P = 179,965^2 - 2
\]

\[
P = 32,387,401,223
\]

Since \(P\) ends with 3, Conjecture 4 is verified.

To confirm Conjecture 5, \(P \pm 2\) must not be primes.

\[
P = 32,387,401,223
\]

\[
P + 2 = 32,387,401,225
\]

\[
P + 2\] is not a prime since it is divisible by 5. Also, it is a perfect square.

\[
P = 32,387,401,223
\]

\[
P - 2 = 32,387,401,221
\]

Also, \(P - 2\) is not prime since it has other factors like 179,963 and 179,967. This verifies Conjecture 5.

To verify Conjecture 6, it must be shown that \(P = 32,387,401,223\) is congruent to 3(\text{mod} 4). Since \(32,387,401,220\) is divisible by 4, this implies

\[
32,387,401,220 \equiv 0(\text{mod}4)
\]

\[
32,387,401,220 + 3 \equiv 0(\text{mod}4) + 3
\]

\[
32,387,401,223 \equiv 3(\text{mod}4)
\]

Since:

\[
32,387,401,223 \equiv 3(\text{mod}4). \text{ Conjecture 6 is confirmed.}
\]

To verify Conjecture 7, it must be shown that \(P = 32,387,401,223\) is congruent to 7(\text{mod} 8). Since \(32,387,401,216\) is divisible by 8, this implies

\[
32,387,401,216 \equiv 0(\text{mod}8)
\]

\[
32,387,401,216 + 7 \equiv 0(\text{mod}8) + 7
\]

\[
32,387,401,223 \equiv 7(\text{mod}8)
\]

Since:

\[
32,387,401,223 \equiv 7(\text{mod}8) \text{ Conjecture 7 is verified.}
\]

To verify Conjecture 8, let \(n = 179,967\). Then this implies that \(n^2 - 2 = 32,388,121,087\). This number is divisible by 7, and \(7 \equiv -1(\text{mod}8)\). Conjecture 8 is confirmed.

**Justifications**

**Conjecture 1:** If \(P > 2\) and \(P = n^2 - 2\) is prime, then \(n\) is odd.

**Proof:**

To prove that only an odd number \(n\) can make \(P\) prime, assume otherwise. That is, \(n\) is even. This implies that \(n = 2k\).

\[
P = n^2 - 2
\]

\[
P = (2k)^2 - 2
\]

\[
P = 4k^2 - 2
\]

\[
P = 2(2k^2 - 1)
\]

This implies that \(P\) is even, but any even number greater than 2 will never be a prime. A contradiction. **QED.**

**Conjecture 2:** If \(P = n^2 - 2\) is prime and \(n \neq 3\), then \(n \neq \pm 3(\text{mod}7)\).
Proof:

By Conjecture 1, \( n > 2 \) can only be odd. \( N > 3 \) cannot be congruent to \( 3 \) because

\[
\begin{align*}
  n &\equiv 3 \pmod{7} \\
  n^2 &\equiv 9 \pmod{7} \\
  n^2 &\equiv 2 \pmod{7} \\
  n^2 - 2 &\equiv 2 \pmod{7} - 2 \\
  n^2 - 2 &\equiv 0 \pmod{7} \\
  P &\equiv 0 \pmod{7}
\end{align*}
\]

This will imply that \( P \) is divisible by 7; a contradiction when \( P > 7 \). In the same line of argument, \( n \) cannot be congruent to \( -3 \pmod{7} \)

\[
\begin{align*}
  n &\equiv -3 \pmod{7} \\
  n &\equiv 4 \pmod{7} \\
  n^2 &\equiv 16 \pmod{7} \\
  n^2 &\equiv 2 \pmod{7} \\
  n^2 - 2 &\equiv 2 \pmod{7} - 2 \\
  n^2 - 2 &\equiv 0 \pmod{7} \\
  P &\equiv 0 \pmod{7}
\end{align*}
\]

Therefore \( n \) cannot be congruent to \( \pm 3 \pmod{7} \). QED.

Conjecture 3: If \( P = n^2 - 2 \) is prime, then \( n \not\equiv \pm 6 \pmod{17} \)

Proof:

Using the same argument in Conjecture 2, \( n \) cannot be congruent to \( 6 \pmod{17} \) because

\[
\begin{align*}
  n &\equiv 6 \pmod{17} \\
  n^2 &\equiv 36 \pmod{17} \\
  n^2 &\equiv 2 \pmod{17} \\
  n^2 - 2 &\equiv 2 \pmod{17} - 2 \\
  n^2 - 2 &\equiv 0 \pmod{17} \\
  P &\equiv 0 \pmod{17}
\end{align*}
\]

This will imply that \( P \) is divisible by 17; a contradiction. In the same line of argument, \( n \) cannot be congruent to \( -6 \pmod{7} \)

\[
\begin{align*}
  n &\equiv -6 \pmod{17} \\
  n &\equiv 11 \pmod{17} \\
  n^2 &\equiv 121 \pmod{17} \\
  n^2 &\equiv 2 \pmod{17} \\
  n^2 - 2 &\equiv 2 \pmod{17} - 2 \\
  n^2 - 2 &\equiv 0 \pmod{17} \\
  P &\equiv 0 \pmod{17}
\end{align*}
\]

Therefore \( n \) cannot be congruent to \( \pm 6 \pmod{17} \). QED.

Conjecture 4: Let \( P = n^2 - 2 \) be prime. If \( P > 2 \), then \( P \equiv d \pmod{10} \) where \( d \in \{3, 7, 9\} \)

Proof:

By the previous conjecture, only odd \( n \) can make \( P \) prime. Thus \( n \) can only be 1, 3, 5, 7, or 9 \( \pmod{10} \).

Case 1: \( n \equiv 1 \pmod{10} \)

\[
\begin{align*}
  n &\equiv 1 \pmod{10} \\
  n^2 &\equiv 1 \pmod{10}
\end{align*}
\]
\[ n^2 - 2 \equiv 1(mod\,10) - 2 \]
\[ n^2 - 2 \equiv -1(mod\,10) \]
\[ n^2 - 2 \equiv 9(mod\,10) \]
\[ P \equiv 9(mod\,10) \]

So \( P \) can end with 9.

Case 2: \( n \equiv 3(mod\,10) \)
\[ n \equiv 3(mod\,10) \]
\[ n^2 \equiv 9(mod\,10) \]
\[ n^2 - 2 \equiv 9(mod\,10) - 2 \]
\[ n^2 - 2 \equiv 7(mod\,10) \]
\[ P \equiv 7(mod\,10) \]

So \( P \) can end with 7.

Case 3: \( n \equiv 5(mod\,10) \)
\[ n \equiv 5(mod\,10) \]
\[ n^2 \equiv 25(mod\,10) \]
\[ n^2 \equiv 5(mod\,10) \]
\[ n^2 - 2 \equiv 5(mod\,10) - 2 \]
\[ n^2 - 2 \equiv 3(mod\,10) \]
\[ P \equiv 3(mod\,10) \]

So \( P \) can end with 3.

Case 4: \( n \equiv 7(mod\,10) \)
\[ n \equiv 7(mod\,10) \]
\[ n^2 \equiv 49(mod\,10) \]
\[ n^2 \equiv 9(mod\,10) \]
\[ n^2 - 2 \equiv 9(mod\,10) - 2 \]
\[ n^2 - 2 \equiv 7(mod\,10) \]
\[ P \equiv 7(mod\,10) \]

So \( P \) can end with 7.

Case 5: \( n \equiv 9(mod\,10) \)
\[ n \equiv 9(mod\,10) \]
\[ n^2 \equiv 81(mod\,10) \]
\[ n^2 \equiv 1(mod\,10) \]
\[ n^2 - 2 \equiv 1(mod\,10) - 2 \]
\[ n^2 - 2 \equiv 9(mod\,10) \]
\[ P \equiv 9(mod\,10) \]

Therefore, \( P \) will only end with 3, 7 or 9.

**QED.**

**Conjecture 5:** If \( P = n^2 - 2 \) is prime and \( P \neq 7 \), then has no twin prime.

**Proof:**

Twin primes are primes that have a difference of two. Some twin primes are 3 and 5, 5 and 7, 11 and 13, etc.

There could be two possible cases: either \( P \) is the lesser twin prime or \( P \) is the greater twin prime. It must be shown that in both cases, \( P \) must have no twin prime.

Case 1: \( P \) is the lesser twin prime

Since \( P \) is the lesser twin prime, then \( P + 2 \) must be checked if it is indeed a prime.

\[ P = n^2 - 2 \]
\[ P + 2 = (n^2 - 2) + 2 \]
\[ P + 2 = n^2 \]

Since \( P + 2 \) is a perfect square, then it cannot be a prime.

Case 2: \( P \) is the greater twin prime

Since \( P \) is the greater twin prime, then \( P - 2 \) must be checked if it is indeed a prime.

\[ P = n^2 - 2 \]

\[ P - 2 = (n^2 - 2) - 2 \]

\[ P - 2 = n^2 - 4 \]

Since \( P - 2 \) has two other factors, then it cannot be a prime. Note that \( n - 2 \neq 1 \) because \( P \neq 7 \). QED.

Conjecture 6: If \( P = n^2 - 2 \) is prime and \( P > 2 \), then \( P \equiv 3(mod 4) \)

Proof:

Since \( n \) is odd (by Conjecture 1), there are only two possible cases: \( n \equiv 1(mod 4) \) or \( n \equiv 3(mod 4) \)

Case 1: \( n \equiv 1(mod 4) \)

\[ n \equiv 1(mod 4) \]

\[ n^2 \equiv 1(mod 4) \]

\[ n^2 - 2 \equiv 1(mod 4) - 2 \]

\[ n^2 - 2 \equiv 3(mod 4) \]

\[ P \equiv 3(mod 4) \]

So \( P \) is congruent to \( 3(mod 4) \)

Case 2: \( n \equiv 3(mod 4) \)

\[ n \equiv 3(mod 4) \]

\[ n^2 \equiv 9(mod 4) \]

\[ n^2 - 2 \equiv 1(mod 4) - 2 \]

\[ n^2 - 2 \equiv 3(mod 4) \]

\[ P \equiv 3(mod 4) \]

So \( P \) is congruent to \( 3(mod 4) \). QED.

Conjecture 7: If \( P = n^2 - 2 \) is prime and \( P > 2 \), then \( P \equiv 7(mod 8) \)

Proof:

Since \( n \) is odd (by Conjecture 1), there are only four possible cases

\[ n \equiv 1, 3, 5, or 7(mod 8). \]

Case 1: \( n \equiv 1(mod 8) \)

\[ n \equiv 1(mod 8) \]

\[ n^2 \equiv 1(mod 8) \]

\[ n^2 - 2 \equiv 1(mod 8) - 2 \]

\[ n^2 - 2 \equiv 7(mod 8) \]

\[ P \equiv 7(mod 8) \]

So \( P \) is congruent to \( 7(mod 8) \)

Case 2: \( n \equiv 3(mod 8) \)

\[ n \equiv 3(mod 8) \]

\[ n^2 \equiv 9(mod 8) \]

\[ n^2 - 2 \equiv 1(mod 8) - 2 \]

\[ n^2 - 2 \equiv 9(mod 8) \]
\[ n^2 - 2 \equiv 1 \pmod{8} - 2 \]
\[ n^2 - 2 \equiv 7 \pmod{8} \]
\[ P \equiv 7 \pmod{8} \]

So \( P \) is congruent to \( 7 \pmod{8} \)

Case 3: \( n \equiv 5 \pmod{8} \)
\[ n \equiv 5 \pmod{8} \]
\[ n^2 \equiv 25 \pmod{8} \]
\[ n^2 \equiv 1 \pmod{8} \]
\[ n^2 - 2 \equiv 1 \pmod{8} - 2 \]
\[ n^2 - 2 \equiv 7 \pmod{8} \]
\[ P \equiv 7 \pmod{8} \]

So \( P \) is congruent to \( 7 \pmod{8} \)

Case 4: \( n \equiv 7 \pmod{8} \)
\[ n \equiv 7 \pmod{8} \]
\[ n^2 \equiv 49 \pmod{8} \]
\[ n^2 \equiv 1 \pmod{8} \]
\[ n^2 - 2 \equiv 1 \pmod{8} - 2 \]
\[ n^2 - 2 \equiv 7 \pmod{8} \]
\[ P \equiv 7 \pmod{8} \]

So \( P \) is congruent to \( 7 \pmod{8} \)

Therefore, \( P \equiv 7 \pmod{8} \). \( QED. \)

Conjecture 8: If \( n^2 - 2 \) is not prime, then its prime factor \( p \equiv \pm 1 \pmod{8} \)

Proof:

All primes are congruent to 1, 3, 5, or 7 (mod 8). It is necessary to separate what kind of primes divide non-prime \( n^2 - 2 \).

Since it is assumed that \( n^2 - 2 \) is not prime, then if follows
\[ n^2 - 2 \equiv 0 \pmod{p} \]

For some prime \( p \).

Now,
\[ n^2 - 2 \equiv 0 \pmod{p} \]
\[ n^2 \equiv 2 \pmod{p} \]
\[ 2 \equiv n^2 \pmod{p} \]

Thus, 2 is a quadratic residue of \( p \). Therefore, \((\frac{2}{p}) = 1\).

The second supplement to the law of quadratic reciprocity (Landau, 1966) asserts that
\[ (\frac{2}{p}) = (-1)^{\frac{p^2-1}{8}} \]

Since \((\frac{2}{p}) = 1\) and only values of \( p \) that makes \( \frac{p^2-1}{8} \) even will be acceptable. Let \( p = 8x + a \). If \( a \) is even, then \( p^2 - 1 \) is odd. So, \( a \) must be odd. Checking for different values of \( a \):
\[ \frac{(8x+1)^2-1}{8} = 8x^2 + 2x \quad \text{even} \]
\[ \frac{(8x+3)^2-1}{8} = 8x^2 + 6x + 1 \quad \text{odd} \]
\[ \frac{(8x+5)^2-1}{8} = 8x^2 + 10x + 3 \quad \text{odd} \]
\[ \frac{(8x+7)^2-1}{8} = 8x^2 + 14x + 6 \quad \text{even} \]
These shows \(a\) can only be 1 or 8.

That is only
\[
p \equiv \pm 1 \pmod{8}
\]
can make \((-1)^{\frac{p^2-1}{8}} = 1\), therefore all prime factors of non-prime \(n^2 - 2\) are congruent to \(\pm 1 \pmod{8}\). QED.

Summary

In summary, this mathematical investigation shows that prime numbers of the form \(P = n^2 - 2\) can only have odd \(n\) (with the exception of \(n = 2\)). \(n\) cannot be congruent to \(\pm 3 \pmod{7}\) (with the exception of \(n = 3\)), and \(n\) cannot be congruent to \(\pm 6 \pmod{17}\).

This also shows that all primes of the form \(P = n^2 - 2\) that are greater than 2 can only end with 3, 7, or 9. In addition, all primes of the form \(P = n^2 - 2\) will never have a twin prime (with the exception of \(P = 7\)). All primes of the form \(P = n^2 - 2\) that are greater than 2 are congruent to 3(mod4) and to 7(mod8). Lastly, all prime factors of non-prime \(n^2 - 2\) are congruent to \(\pm 1\) modulo 8.

Suppose one is finding values of \(n\) that will make \(P = n^2 - 2\) prime. Let \(q\) be a natural number, and \(N_q\) be the set of natural numbers less than \(q\).

Since \(n\) can only be odd, only approximately \(\frac{1}{2}\) of \(N_q\) will be considered. Also, only approximately \(\frac{5}{7}\) of these odd numbers cannot be congruent to \(\pm 3 \pmod{7}\). In addition, approximately \(\frac{15}{17}\) of these odd numbers cannot be congruent to \(\pm 6 \pmod{17}\). This implies that only approximately \(\frac{1}{2} \cdot \frac{5}{7} \cdot \frac{15}{17} = \frac{75}{238} \approx \frac{1}{3}\) of \(N_q\) can be the possible values of \(n\).

Possible Extensions

By Conjecture 7, \(P \equiv 7 \pmod{8}\). This implies that \(P\) can be expressed in the form \(8k - 1\). Now, what are the properties of \(k\) that will make \(8k - 1\) prime, and at the same time, with the form \(n^2 - 2\)?

References


Friends, Birthdays and Probability: Problems Concerning the Probability of Birthday Coincidences?

Marco Lee Solano
Trinity Grammar School, Summer Hill NSW, Australia
E-mail: marco.leesolano@gmail.com

Abstract
I have always been fascinated with coincidences. Viewing events in the world through the scope of chance and probability has become a habit, which may go back to my interest in Pokémon as a child. Pokémon is a game in which a major objective is “catching” an animal-type thing, which is greatly reliant on chance: a certain probability is associated with catching each animal on the first try, and this is often very unlikely. This means that a player must anticipate how many tries they will need to catch the animal (the Pokémon) because attempts must be purchased in advanced. A particular coincidence that was a source of constant bewilderment was that in my class of 31 boys (from year 8 to 10), there were two boys, Andrew and Adrian, who shared their birthday. I thought, “What are the chances of that”? We thought that this was a remarkable coincidence, since it seemed very unlikely and impressive. Since then, I have been introduced to the concept of the Birthday Problem, also called the birthday paradox. This math problem is attributed to Richard von Mises, who first mentioned it in 1939 [1]. This problem is not a paradox in the traditional meaning of a contradiction in logic, however the solution is counterintuitive to some, and I was tricked the first time. The original question asks “how many people are needed in a room such that the chance that some share a birthday is at least 50%?” The well-known answer to this question is 23. People often talk about the “significance”, or “size” of a coincidence. One might say that two people tossing coins with the same result was a relatively small coincidence, whereas meeting someone with the same birthday as you would be a relatively large coincidence. It seems that the “significance”, or the “size” of a coincidence is an expression for how unlikely a certain coincidence is to occur. I have always wondered what the chances of our classroom birthday coincidence occurring were, and how many friends you would usually need to have before you saw a coincidence like that.

Introduction
Objective 1: To investigate the chances of my classroom birthday coincidence occurring. Objective 2: To determine how many friends I need so that two (at least) among us share birthdays.

When considering the chances of my classroom coincidence occurring, we need to consider the circumstances: there was a pair of boys sharing a birthday out of a classroom of 31 children. One way of approaching this question is to determine the probability of any birthday coincidence occurring. This question relates closely to the original Birthday Problem, or Birthday Paradox, which asks “how many people do you need in a room for there to be a 50% chance of two or more people sharing a birthday?”

The problem posed above can be solved with the assumption that there are 365 days in a year, and that a person is equally likely to be born on any day of the year [2, 3].

Let \( P(n) \rightarrow \) “the probability of two or more people sharing a birthday in a group of size \( n \) \( (n \in Z^+) \)

A problem arises in the consideration of this questions since as values of \( n \)
increases there are increasing numbers of types of events to consider. For a group of \( n > 2 \) there are multiple ways coincidences can occur? For example, with \( n = 3 \) people, there is also the possibility of a three-way coincidence (3 people sharing a single birthday), with \( n = 4 \) people, there is the possibility of a single two way coincidence, two way coincidences, a three way coincidence, or a four way coincidence. As is evident, this quickly gets out of hand. Consequently, I decided a different approach was appropriate.

**Process of Derivation: Approaching the Complementary Event**

I tried approaching the problem by finding an expression for a coincidence not occurring. In other words considering the complementary event appeared easier than a direct approach, so I attempted to determine the probability of \( P' \) (the complementary event of \( P \)). I started with smaller values of \( n \) to try to derive an equation.

\[
P'(1) = 1
\]

Since there is no one for the lone person to share a birthday with.

\[
P'(2) = 1 \times \frac{364}{365}
\]

Since there are 364 possible remaining days out of the 365 for the second person to be born on without a coincidence occurring

\[
P'(3) = 1 \times \frac{364}{365} \times \frac{363}{365}
\]

Since, with two people already not sharing birthdays, there remain 363 possible remaining days out of the 365 for the third person to be born on without a coincidence occurring. This can be expressed as:

\[
P'(3) = \frac{365}{365} \times \frac{365 - 1}{365} \times \frac{365 - 2}{365}
\]

I noticed that this was a recurring pattern, and can be written as, since each additional person in the group of \( n \) will have one \( 365 - (n - 1) \) birthdays left to choose from if they are to not share a birthday with anyone.

For any value of \( n \),

\[
P'(n) = \frac{365}{365} \times \frac{365 - 1}{365} \times \frac{365 - 2}{365} \times \ldots \times \frac{365 - (n - 1)}{365}
\]

To refine this expression:

- The numerator resembles a factorial function. A factorial function is defined as follows where \((n!)\) means the product of all the positive integers up to and including \( n \).
- I recognized that the numerator could be expressed in a similar way by the way that the numerator of each successive fraction that is multiplied in the equation is one less than the one preceding it (365, 364, 363…), which is a little bit like 365!
- However, we notice that the numbers in the numerators do not include integers that are lower than \( 365 - (n - 1) \). This is called a partial factorial through research [4]. I found that the numerator could be expressed by the following expression:
Having studied conditional probability in Year 10, I recognized \( \frac{365!}{365-(n-1)!} \) as being equal to \(^nP_n\) the number of \( n \) permutations of 365. This makes sense since it represents the number of ways that \( n \) different birthdays can be selected from the possible 365. \( xP_n \) signifies a permutation function [5], also denoted by \( xP_n \) and \( P(X,Y) \), where the value is defined by:

\[
xP_n = \frac{X!}{(X-Y)!}
\]

In addition, the denominator can be simplified to \( 365^n \) which makes senses since this is the total number of possible ways that \( n \) people can have their birthdays. The first person can have their birthday any of 365 ways, the second person can have their birthday any 365 ways, etc.

To include these simplifications, the final expression for the probability of no birthday coincidence occurring in a room of \( n \) people is

Equation 1: \( P'(n) = \frac{365P_n}{365^n} \)

And therefore, where \( P(n) \rightarrow \) “the probability of two or more people sharing a birthday in a group of \( n \) people”:

Equation 2: \( P(n) = 1 - \frac{365P_n}{365^n} \)

Findings: Objective 1

To solve this for the number of people in our class I substituted 31 as the value for \( n \), and used the computational knowledge engine Wolfram Alpha to solve for \( P \). This method was used since numbers such as “365!” And “36531” are far too large for even our graphics calculators to compute. They return errors every time. The printout of Wolfram Alpha:

\[
P(31) = 1 - \frac{365P_{31}}{365^{31}}
\]

\[
P(31) = 0.730455
\]

Therefore, the probability of a coincidence in a class of 31 people is approximately 73% meaning there was only 27% chance of there not being a coincidence. I have graphically displayed the values of \( P(n) \) with \( n \) up to \( n = 120 \) people. The red data point shows the probability of a coincidence when there are 31 people in a room.
Figure 1: A graph of the probability of a birthday coincidence occurring groups of size $n$ for values of $n$ up to 120.

**Reflection: Objective 1**

73% is disappointingly high, since I had been hoping to find that our coincidence was rare and noteworthy. It seems that there was nothing remarkable about the coincidence in our class. What surprised me the most about these results is that there was in fact a far greater chance of a coincidence occurring than one NOT occurring? It would be more statistically remarkable if our class did NOT have any coincidences. This suggests that the occurrence of a birthday coincidence in our class is not significant at all.

This sparked my interest since it illustrates that seemingly improbable events related to coincidences of random variables may be far more common than we expect, even when the variable can have 365 values. The calculated result demonstrates that birthday coincidences are actually very common, despite our class’ original assumption that the event was improbable. I personally believe that the reason people overestimate the significance of a birthday coincidence is that the chances of a person sharing your birthday are so much lower than this, since about 1/365 people you meet will have the same birthday as you. While this is merely personal conjecture, it may be that since people are likely to know very few people who have the same birthday as them, and thus overestimate the frequency of birthday coincidences in general. This is an area of further research potentially in the field of psychology, which may explain why many people are surprised by the frequency of birthday coincidences. An investigation into the probabilities of people sharing your own birthday could add to what I have learnt through this project.

A second important aspect of this result is that the equations I derived are generalizable to determine the probability of a coincidence in a group of people any size, but even to other situations involving a discrete random variable that has a set number of values. Equation 2 can be adapted to calculate the probability that a discrete random variable with a set number of $l$ possible values will repeat a value it has already taken in a previous trial, given $n$ trials.

$$P(n) = 1 - \frac{iP_n}{l^n}$$

An example of this would be in calculating the probability of rolling a number of dice and having 2 or more give the same value. This has implications for chance based games, such as Yahtzee, where repeating a previously obtained dice roll is favourable to the player. Also, coincidence of random variables is an integral part of the Pokémon game, where in each attempt that a player makes to “catch” a Pokémon, the game generates 3 random numbers. These numbers must coincide for the Pokémon to be caught, and different Pokémon have different “catch rates”, meaning that the range of possible numbers generated is smaller or larger, increasing and decreasing (respectively) the likelihood of the player catching the Pokémon [5]. In addition, this result has some limitations. My methods had many accompanying assumptions, such as the equal probability of birthdays.
occurring on every day of the year, and the exclusion of leap years. While these may not have had a major or even noticeable effect on the final result, the methods I used were not able to take these effects into account, and this was the most significant limitation of the investigation so far.

Even so, when reflecting on my result, it doesn’t exactly answer the original question of “how likely is that?” The approximate 73% chance that I calculated is the chance of one or more birthday coincidences occurring. Technically we didn’t just have any coincidence, Adrian and Andrew shared birthdays, but everyone else had different birthdays. Therefore, even though I was able to devise a sound method of determining the probability of any birthday coincidence occurring in my year 8 class, I did not succeed in entirely completing objective 1 of my aim. Hence, I decided to use conditional probability to determine the probability of only 2 people sharing a birthday in a class of 31 people.

To Determine the Probability of Only Two People Sharing A Birthday

After having derived a formula to solve the initial birthday paradox, I found the second much easier, and I was able to logically deduce a formula. This was the process of thinking I followed when deriving an equation for the probability of only 2 people sharing a birthday in a room of n people.

Process of Derivation: A Specific Arrangement of Birthdays

If two people share a birthday in a room of n people, there are:

- n – 1 Occupied birthdays. One property of \( x^P_y \) permutation is that it gives the number of ways that a sequence of y items can be selected in a “sequence without repetition” from a set of x items [4]. Therefore \( 365^P_{n-1} \) ways for n – 1 birthdays to be occupied out of 365 possible days.

- Two people sharing a birthday. We must determine the number of different possible pairs of people that can be arranged in our room of size x. From year 10 maths, I recognized this as being a problem that required combinations. The notation \( xC_y \) signifies a Combination function, also denoted by \( xC_y \) and \( C(x,y) \) where the value is defined by:

\[
x \cdot C_y = \left[ \frac{X!}{Y! \times (X - Y)!} \right]
\]

\( xC_y \) is also equal to the number of possible combinations of y items that you can select from a larger set of x (where order does not matter) [6].

Since we are looking for the number of different pairs you can select from the set of n people. Therefore \( C_2 \) ways that 2 people can be selected from the number of people who are in the room of n.

This represents the number of ways that y items can be selected from a set of x with replacement, and where the order does not matter. Therefore the number of ways that
\(n\) people can have birthdays such that only two of them share a birthday is

\[\binom{n}{2} \times 365^{P_{n-1}}\]

Therefore the probability of only two people in a room of \(n\) people sharing a birthday is the number of ways that it can happen, \(\binom{n}{2} \times 365^{P_{n-1}}\) divided by the total number of ways that \(n\) people could have birthdays, 365\(^n\).

Equation 3: \(P_{\text{single pair}} = \frac{\binom{n}{2} \times 365^{P_{n-1}}}{365^n}\)

Where, \(P_{\text{single pair}}\) is the probability of only two people sharing birthday. Substituting 31 as a value for \(n\)

\[P_{\text{single pair}} = \frac{31 \times 365^{30}}{365^{31}}\]

**Findings: Objective 1 (Second Attempt)**

Using Wolfram Alpha again, I derived the following result, by solving equation 3 for \(n = 31\)

\[P_{\text{single pair}} = 0.3741\ (4\ s.f.)\]

Graph 2: A graph of values (up to \(n = 100\) people) of the probability of there being ONLY 2 people sharing a birthday in a group of \(n\) people with respect to \(n\) is shown.

**Reflection: Objective 1 (Second Attempt)**

Evidently the probability of a class such as ours to have exactly one pair of people sharing a birthday is 37\% (highlighted by the red data point). This is still a surprisingly high chance, since the finding implies little significance to the coincidence of our class. The chance of only two people sharing their birthday in a class of 31 is still greater than the chance of no people sharing birthdays 37\% versus 27\%). More interestingly however, closer inspection of the trend of this function reveals that a room or 31 people yields close to the highest chance of only having one birthday-sharing-pair. The highest chance of a single only two people sharing birthdays occurs in a group of 28 people (38.6\% chance), and for any more people the chance decreases. It makes sense that the likelihood of this decrease as the number of people in your room increases (in a room of 300+ people, it
would intuitively seem very unlikely that 
only a single pair of people share a 
birthday).

It is very satisfying to have finally resolved 
this question that I have wondered about 
since middle school, and telling my friends 
Andrew and Adrian that their coincidence 
has a high chance of occurring in all 
classroom settings surprised them as well. 
With these results, I was eager to test how 
true my predictions were. In order to do 
this, I went to a random number generator 
(www.random.org) and set it to generate 10 
sets of 31 random numbers ranging from 
1–365 inclusive, mimicking the 
randomness of birthdays. I then analysed 
these sets. I found that 4 out of the 10 sets 
had just 2 of the same number, n concordance with my calculated prediction 
of a 37% probability of ONLY 2 people 
having the same birthday (4 / 10 ≈ 37%). 
Furthermore, I found that 7 out of the 10 
sets had at least 1 pair of coinciding 
numbers, in concordance with my 
calculation from my first attempt at 
objective 1, where I discovered that the 
probability of a birthday coincidence 
occurring in a set of 31 people was 73% (7 / 
10 ≈ 4 / 10).

**Objective 2: To Determine How Many Friends I need so that two (At 
Least) among us share Birthdays**

Having resolved objective 1 with surprising 
results, I was then faced with the question of 
“how many friends do I NEED so that 
among us there will be a shared birthday?” 
This problem was actually much easier to 
answer than I had anticipated. According to 
the pigeonhole principle, “if n pigeons are 
placed into m pigeonholes, and n > m, , 
there must exist at least one pigeonhole with 
2 pigeons in it”.

The same applies for birthdays: if I assign 
366 people to birthdays in the year, then 
there must be at least 2 people on one 
birthday since there are only 365 possible 
birthdays. It would still be possible for 365 
people to not have coinciding birthdays, 
although this is ridiculously unlikely.

Therefore, the answer is **366 friends**. This 
answer was unsatisfying though. The point 
of the birthday problem is that it shows that 
not very many people in a room are 
required for a high probability of a birthday 
being shared. So let’s modify this question: 
how many friends (including myself) do I 
need ON AVERAGE for there to be a 
birthday coincidence?

**Objective 2 (Revised): To Determine How Many Friends I need on 
Average so that there is a Birthday 
Coincidence**

This is much more difficult to answer. At 
first I assumed that since “number of 
people” is a discrete variable, I would be 
able to use a formula such as this [7]:

\[
E[n] = n_1p_1 + n_2p_2 + \ldots + n_kp_k
\]

- where \( E[n] \) is the expected value of \( x \) 
  (weighted average of the terms), and 
- \( n_kp_k \) is an event \( (n) \) multiplied by the 
  chance of that occurring \( (p) \).
- when there are \( k \) values of \( n \).
This could also be written as [7]

\[ E[n] = \sum_{n=1}^{k} n \times p(n) \]

- where \( p(n) \) is the probability of the event \( n \) occurring
- where \( n \) can only have integer values from 1 to \( k \)

However, this cannot be applied to the results that I have derived so far. This is because a birthday coincidence in a group sized \( n \) (which Equation 2 describes the probability of) is not mutually exclusive to the event of a birthday coincidence occurring in a group of people other than \( n \). This equation I have for \( E(n) \) requires that the events are mutually exclusive and collectively exhaustive [7]. Mutually exclusive events cannot happen at the same time. Collectively exhaustive means there are no other possibilities other than the events in questions, and that the sum of the probabilities of these events occurring is 1 \( (p_1 + p_2 + p_3 + \ldots + p_k = 1) \) that which evidently is not the case. This is clear when considering even low values of \( n \). The probability of a birthday coincidence occurring in a group of 40 people:

\[ P(40) = 1 - \frac{365^40}{365^{40}} = 0.89123 \]

The probability of a birthday coincidence occurring in a group of 41 people:

\[ P(41) = 1 - \frac{365^{41}}{365^{41}} = 0.90315 \]

\[ 0.89123 + 0.90315 > 1 \]

This is because a group of 30 people has within it groups of 29 people, 28 people etc. This means that when I randomly select a group of \( n \) birthday, there are groups of \( n-1, n-2, n-3, \ldots \) people that have been selected within that group. This means that the event of \( n \) birthdays being selected with a birthday coincidence can include the event of a smaller group of birthdays being selected with a birthday coincidence: they can happen at the same time, therefore they are not mutually exclusive. Evidently, I either needed a different method of calculating averages, or I needed to derive a new formula. I had never encountered a probability question such as this, where I could not conceptualize a way of expressing the events in a way that made them mutually exclusive.

**Alternative Perspective: Repeated Selections of Items from a Set**

For a great deal of time, I could not deduce a way of solving this problem. I would require a new way of considering groups of people with birthdays. What I eventually came to notice was that observing a group \( n \) of people (whose birthdays we have assumed to be randomly allocated) was akin to randomly selecting a set of \( n \) numbers from a set of 1 to 365.

This has an important implication for the nature of this problem: if the birthdays of a group of \( n \) people can be represented by a set of \( n \) random numbers selected from the set of [1, 365], then it is also akin to selecting \( n \) random numbers from the [1, 365] one at a time with selection. Therefore, the probability of a birthday coincidence occurring in a group of \( n \) people is the same.
as the probability of getting the same number(s) twice when you select \( n \) random numbers in succession with replacement from the set of \([1, 365]\).

This gave me the opportunity to consider how the probability of a reselection of a number occurring in a group of size \( n \) can be mutually exclusive for other group sizes. Now however, it is easy to think of the question as “if I make successive selections of numbers from the set \([1, 365]\), how many times do I have to retry on average for the selection of a previously selected number to occur?”

This question involves the calculation of how likely the \( n \)th trial is to produce the first reselection. As I thought about this, I realized that for this event to occur, two conditions must be satisfied.

For the \( n \)th trial to be the first to select an item that has previously selected:

All trials previous to the \( n \)th must have failed: no reselection can have occurred in any previous trials.

And,

The \( n \)th trial must select an item that has already been selected.

Since these conditions must co-occur if the trial is to succeed, a general expression would be \((\text{Probability of no previous reselections}) \times (\text{Probability of a reselection occurring in this trial})\)

In order to determine the efficacy of this method, I tried applying this to a smaller scale population of 4 items.

### Investigating a Smaller Case

To better understand this problem, I took a much smaller population to investigate. Let’s consider four different items in a bag, denoted by A, B, C and D. How many times do I have to select (with replacement) out of the bag on average before I pick an item I have already chosen? In accordance with the pigeonhole principle, it is evident that I will have to make 5 random selections at most to pick an item twice. Also, the chance of a reselection after I have made 1 selection is 0. Therefore the answer to this problem lies between 1 and 5.

I have calculated the probability of each trial from 1 to 5 producing the first successful reselection, by calculating the probabilities of the two required conditions for success.

What is the chance of my first reselection of an item being on the FIRST trial?
- The chance that no item has been reselected already is 1. It is certain.
- The chance that I select an item that has already been selected in a previous trial is 0. There were no previous trials.
  Therefore the chance is \( 1 \times 0 = 0 \)

What is the chance of my first reselection of an item being on the SECOND trial?
- The chance that no item has been reselected in a previous trial is 1.
- I have selected one item already, and there are four items that I could
pick. Therefore the chance of success in the first trial is

$$1 \times \frac{1}{4} = \frac{1}{4}$$

What is the chance of my first reselection of an item being on the THIRD trial?
- The chance of no previous trials reselecting an item is $\frac{3}{4}$ because one item had been selected in the first trial, and $\frac{3}{4}$ is the probability of any other item being selected.
- Since 2 items have been selected in the previous trials, the chance of successful reselection of one of these in this trial is $\frac{2}{4}$. Therefore the chance of my first reselection of an item being on the THIRD trial is $\frac{3}{4} \times \frac{2}{4} = \frac{3}{8}$.

What is the chance of my first reselection of an item being on the FOURTH trial?
- This requires that in all previous trials, there were no reselections; i.e. three different items were selected beforehand.
- Just like **Equation 1**, the probability of this is $\frac{4}{4} \times \frac{3}{4} \times \frac{2}{4}$.
- With three items previously selected, the chance of reselecting one on the fourth trial is $\frac{3}{4}$.

Therefore the chance of my first reselection of an item being on the FOURTH trial is

$$\frac{3}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{9}{32}$$

What is the chance of my first reselection being on the FIFTH trial?
- The chance of NO reselections over the previous four trials is $\frac{4}{4} \times \frac{3}{4} \times \frac{2}{4} \times \frac{1}{4}$.
- The chance of selecting an item that has already been selected is 1: all four items have already been selected once in the previous trials. Therefore the chance of my first reselection of an item being on the FIFTH trial is $1 \times \frac{3}{4} \times \frac{2}{4} \times \frac{1}{4} = \frac{3}{32}$.

What is the chance of my first reselection being on any subsequent trial after the 5th?
- The chance of no reselections over the previous four trials is 0.
- The chance of an item being reselected that has already been selected doesn’t matter, because anything multiplied by 0 is 0. Therefore the chance of my first reselection of an item being on a subsequent trial is 0.

**Application to the Coincidence of Birthdays**

We remember from the original birthday paradox solutions, that the expression for the probability of no birthday coincidence occurring in a room of $n$ people is given by:
Equation 1: \[ P'(n) = \frac{365 P_n}{365^n} \]

“No previous reselections” means there is no reselection of an item occurs in the trials before the trial in question. In order to make the expression more general, I will write the equations with a population size of “I” different numbers, instead of 365 (which would correspond to possible birthdays).

Therefore in the \( n \)th trial of randomly selecting items from a total population of \( l \) the probability of all trials preceding the \( n \)th having no reselections (coincidences) is:

Equation 4: \[ \frac{l P_{n-1}}{l^{n-1}} \]

This is the probability of \( (n-1) \) items being chosen from a population of \( l \) without any coincidence occurring.

Hence, on the \( n \)th trial the chance of a reselection occurring, since \( (n-1) \) items have been previously selected out of a population of \( l \), is:

Equation 5: \[ \frac{n-1}{l} \]

Therefore, if I randomly make successive selections of items from a group of \( l \) items, the chance of the \( n \)th selection being the first time any item was selected for a second time is:

(Probability of no previous reselections) \( \times \) (Probability of a reselection occurring in this trial)

Substituting Equation 4 and Equation 5:

\[ \frac{1}{l^n-1} \times \frac{n-1}{l} \]

\[ \frac{(n-1)}{l^n} \]

With the question of birthdays, the population, \( l \), is 365 since there are 365 days (“items”) I could select, so the probability of the \( n \)th date I randomly select coinciding with one that I previously select is:

Equation 6: \[ \frac{(n-1)}{365^{n-1}} \]

Findings: Objective 2

As before, I now need to calculate a weighted average for all possible values of \( n \) [8]:

\[ E[n] = \sum_{n=1}^{\infty} n \times p(n) \]

- where \( E[n] \) is the probability of the event \( n \) occurring
- where \( p(n) \) is the probability that \( n \) will occur

Note however that for values of \( n \) that are greater than 366, the probability of that trial producing the first coincidence is 0, because according to the pigeonhole principle, the 366th trial must have produced a coincidence. So a refined expression for my use would be:
The event $n$ in the case of birthday coincidences is the event that the $n^{th}$ trial will produce the first successful reselection of an item that has been previously selected, and therefore $p(n)$ is expressed by Equation 6. Substituting Equation 6 for $p(n)$:

$$E[n] = \sum_{n=1}^{366} n \times p(n)$$

I used a spreadsheet to calculate all values of \left[ n \times \frac{(n-1)^{365}P_{n-1}}{365^n} \right] up to $n = 366$, then add them all together which gave me the answer of $E[n] = 24.61 \ (4. \ s.f.)$.

This is surprisingly small, and suggests that on average, a birthday coincidence occurs in every class of 25 or more children. This was a valuable addition to my understanding of the birthday paradox since it demonstrates the abundance of “coincidences” that are actually very common. I have included a graphical representation of this summation as well, where the red area underneath the distribution is equal to the solution of this sum. Notice how after $n = 27$, the probability begins to decrease. This means that when you make more than 27 trials, the chance of this number producing the first coincidence is lower than that of the trial preceding it. This is because of the probability of previous trials yielding no coincidence becomes very low very quickly, since after $\sim 28$ people, there is a lower and lower chance that there has not already been a coincidence. My findings from objective 1 serve as a testament to the high probability of a coincidence having occurred even with small $n$ such as 23.

Reflection: Objective 2

As my calculations indicated, I need 25 friends on average for there to be a shared birthday between us. The average number of trials necessary for a discreet random variable to repeat itself however has many applications in the real world. One particular example is in computer coding. One process, called Hashing, involves the assignment of data to corresponding numbers using a “Hashing Function”, making the data almost irretrievable without knowing what the original function was that sorted it. However, the pseudo-random nature of the function means that often more than one piece of input data can be assigned to the same number. As I discovered, this happens relatively frequently, even with very large populations. This means that,
should the programmer be Hashing as a means of security for their information, another programmer may be able to determine the nature of the function by randomly inputting values and looking to see when two different pieces input data are assigned to the same hash location [7]. This is a difficult concept for me to understand because I have little experience with coding, however, it is clear that the implications of this mathematical field are far-reaching, in this age of electronic information.

My methods of deriving the equations reflect intuition and many different perspectives that I needed to use. Through reviewing the nature of selection with replacement, I was able to relate this to the question of people’s birthdays coinciding, which meant that I could create a very general solution to the question of the expected number of trials for a discrete random variable to repeat itself in.

This was of significant interest to me since it explained other seemingly unlikely coincidences I had previously been baffled with. One example is that I am a frequenter of short comic websites, such as SMBC and XKCD, both of which have large archives of content and a feature that presents a randomly selected comic. I had noticed however that even without having read many comics, the random feature would often produce a comic that I had previously read which baffled me. I was unable to explain why out of all the many comics that it could produce, I still encountered ones that I had seen before. My findings of 25 friends sufficing for a coincidence out of 365 possible birthdays just served to show that discrete random variables coincide with previous values far more than we would otherwise expect. This only serves to highlight how human intuition can often find it hard to conceptualize large scale probability-related situations, and that rigorous deduction can be used to describe them and better understand these elusive concepts.

The spreadsheet method for calculating the summation was only used as a last resort after I could find no algebraic method for calculating the summation, since (while accurate to within more than 4 decimal places) it is an admittedly inelegant way of calculating the summation. In the following section I will detail some of the methods I employed with failure to try and solve this problem.

**Unsuccessful Attempts at Objective 2: Some Limitations**

For this final problem of the average number of people needed for a birthday coincidence, I tried numerous methods which failed. I attempted a method recommended by mathematician Knuth [9] who describes a method for calculating the probability of a random coincidence for a population size $M$. Knuth was a was a pioneer in the field of random collisions (in coding and computer programming), describing the average number of trials required for a reselection in a population size of $M$ to be:

$$1 + \sum_{n=1}^{M} \frac{M P^n}{M^n}$$

I was not able to understand why a summation of the probabilities of $P'(n) + 1$
was equal to the expected value, but this may be some property of non-mutually exclusive events. However I understand that the “1+…” is because for any \( n \) trials that are conducted until a reselection is made, there were \( n-1 \) trials that no reselection was made, so the average number of trials I need to conduct without a reselection will be exactly one less than the number of trials I need to conduct until I get a reselection. Solving Knuth’s equation with \( M \) as 365 gives the same answer as calculated [10].

My first approach was to try and solve this equation given by Knuth through an integration, when I realized that the answer expressed by that summation was also equal to the area underneath the curve described by this function. However, classical methods of integration required that the function be continuous (factorials and permutations are only defined for integer numbers). So I attempted a number of approaches, first considering the Gamma function as able to express non integer values and make the function continuous, but I found that after I had substituted the appropriate Gamma expressions into Knuth’s function, it was well beyond my means of integration (or any integration calculator I could access). The gamma function is a function that can be used to calculate factorials of numbers. It has the interesting property of being able to give a factorial function of a non-integer value. The Gamma Function is expressed by

\[
\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} \, dx
\]

With the relationship of this to a factorial being:

\[
(t-1)! = \int_0^{\infty} x^{t-1} e^{-x} \, dx
\]

\[
(t-1)! = \Gamma(t)
\]

I read a University lecture handout [11] showing how the integration by parts of the Gamma function led to:

\[
\Gamma(t+1) = t \cdot \Gamma(t)
\]

Which is the defining property of a factorial function:

\[
x! = x \cdot (x-1)!
\]

A second continuous expression for factorials is the Stirling approximation, which while less accurate, could be used as a replacement for the factorials in the equation as follows.

\[
n! \approx \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n
\]

Substituting this into the equation for:

Equation 6: \[
\frac{(n-1) \times 365^n}{365^{n-1}}
\]

(Substituting permutation notations for factorial notations)

\[
= \frac{n \times (n - 1) \times 365!}{365^n \times (365 - (n - 1))}
\]

(Substituting the Stirling approximation for the factorial function in the denominator)

\[
\frac{n \times (n - 1) \times 365!}{\sqrt{2\pi (366 - n)} \cdot \left(\frac{366 - n}{e}\right)^{366-n} \cdot 365^n}
\]
I learned that integrating this was not possible by analytic means and that this was because of the factorial function (!) in the equation.

Taking logs of both sides, the \( \int f(x) \, dx \) was unrecoverable from \( \ln \) and no conventional techniques would allow let calculate this integral (because of the \( n^n \) component, highlighted in yellow [12]).

**Conclusion**

In this investigation I discovered that birthday coincidences, which I had previously thought of as significant unlikely, are actually very frequent, and that in a class of 31 people, it is more likely than not (73% chance of having a coincidence vs 27% chance of NOT having a coincidence) for some children to share a birthday. I learned that the birthdays of my classmates Andrew and Adrian was nothing significant - that it would have been more of a rare occurrence if no one had shared any birthdays. Furthermore, I discovered that the probability of ONLY two people in my class to be a surprisingly high 37% chance. I determined that the average number of friends I need to know for a birthday coincidence to occur in the group is 25.

Therefore, in achieving both my objectives of describing in detail the coincidence of birthdays as well as devising mathematical models for expressing their probability I have satisfied the aims of my investigation.

In a world that is largely governed by chance, my personal search for an explanation of why coincidences occur has taught me that coincidences occur more frequently than we expect in random variables because of the sheer number of coincidences that could occur. I have learnt that often we overestimate the “size” of a coincidence, and that what initially appears to be a significant and unlikely event may be a very probable result of chance.

In the process of attempting to solve this problem, I became familiar with powerful approximations of factorial functions that can determine the factorial of a non-integer number, such as the Gamma function and Stirling approximation. Therefore, in immersing myself in this area of mathematics I have begun to understand far more advanced expressions of the factorial function, something crucial to the area of probability.

The theory behind the coincidence of birthdays also intersects with major aspects of electronic data security and coding, meaning that the programs with which I wrote this paper are also subject to its effects. The high probabilities of coincidences occurring in hashing programs as mentioned earlier pose security threats, which have meant the development of safeguards and more uniform hashing programs.

Furthermore, it is results such as these that allow the true randomness of a set of data to be estimated: if I have a set of data that has too many or too few coincidences, there is a greater likelihood that the data is not random. Equations such as those I devised (equation 6) are often used to determine whether sets of are truly random, and this is important in areas such as scientific simulations, where the simulation is only valid if the data is random [13].
Therefore, the implications of this investigation affect my personal life and perception of the world, but have much more far reaching implications because of the way the methods I used accurately describe the probability of coincidences in discrete random variables, which is a crucial aspect of hashing functions in computing and simulations of data in the sciences. Furthermore, I identified some areas for further research since it is still unclear why the high frequency of birthday coincidences surprises many people.

References


What is the Relationship between the Speed of a falling object and the size and shape of the Crater caused by its impact with the ground?

Andrew Serb
Trinity Grammar School, Summer Hill NSW, Australia
E-mail: aserb.2014@trinity.nsw.edu.au

Abstract
When I read about the meteor that supposedly killed the dinosaurs, I was astonished to find that an object only 10km wide could have caused a crater 193 km wide and 43 km deep, hurling out an estimated 90,000 km³ of debris into the air [1]. After looking at photographs of meteor craters on the moon, I wanted to explore the relationship of an objects speed at impact and the crater it creates. I also wanted to know whether or not a crater was more like a rounded bowl (a cap shape) or a paraboloid. Therefore, in this project, I decided to investigate the relationship between the speed of a falling object and the size and shape of the crater it makes when it impacts the ground. More specifically my aim was to see if there is a relationship between a crater’s volume, circumference and depth when compared to the causative object’s speed. In addition, I wanted to investigate what mathematical formula may best describe the shape of a crater when it hits the Earth. To do this I will set up an experimental rig that will drop a known object of consistent size and weight into a material and measure the crater that it makes. I will also explore whether or not the shape of the crater would change depending on how fast it hits the ground.

Introduction
In order to collect my data and control the variables, I will drop a standard spherical weight (a golf ball) from various heights into a flat, horizontal target zone of sifted, clean, dry sand, in a bucket that is filled to the brim [2].

The diameter and depth of the subsequent crater will be recorded using a digital Vernier callipers (accurate to 0.1mm) for each drop height in centimetres (cm). The crater will then be carefully refilled with sand. The weight of the sand will be measured using tared digital scales (g). By determining the density of the sand (g/cm³) I can calculate the volume of each crater by knowing the weight of sand needed to re-fill the crater. This will be carried out three times for each drop height for mitigation of random error. The drop heights and placement will be standardized through a rig with an adjustable release mechanism. The resultant data will then be displayed graphically, and mathematical correlations evaluated between various parameters using the Pearson’s correlation test. The coefficient of correlation will be found to determine the extent to which speed influences. Finally, the formulas for a cap (part of a sphere), cone and elliptic paraboloid will be used to determine the hypothetical volume of each crater based on its diameter and depth and will be compared graphically and mathematically to the volume data determined by the crater refilling process. These will be compared using percentage differences of the absolute values of the real versus hypothesized volumes to determine which shape profile best correlates to the observed craters.

Information and Measurement
Table 1 shows the raw data for the experiment that was conducted correlating depth, diameter and sand weight displacement with drop height. I have made the assumption that, when dropped from larger heights, the ball will hit the sand at faster speeds. These increments were
chosen because I wanted to record reasonably even and relatively small increments leading up to the tallest height measurable with my tape measure, 2m.

Table 1: Raw Data

<table>
<thead>
<tr>
<th>Drop Height (cm)</th>
<th>Diameter of Crater (cm)</th>
<th>Depth of Crater (cm)</th>
<th>Sand Displaced (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.71</td>
<td>0.83</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>3.55</td>
<td>0.78</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>3.54</td>
<td>0.85</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>4.13</td>
<td>0.98</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>4.19</td>
<td>0.99</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>4.30</td>
<td>1.00</td>
<td>8</td>
</tr>
<tr>
<td>20</td>
<td>4.80</td>
<td>1.18</td>
<td>12</td>
</tr>
<tr>
<td>20</td>
<td>4.83</td>
<td>1.17</td>
<td>11</td>
</tr>
<tr>
<td>20</td>
<td>4.53</td>
<td>1.19</td>
<td>11</td>
</tr>
<tr>
<td>50</td>
<td>5.48</td>
<td>1.60</td>
<td>20</td>
</tr>
<tr>
<td>50</td>
<td>5.27</td>
<td>1.56</td>
<td>19</td>
</tr>
<tr>
<td>50</td>
<td>5.47</td>
<td>1.53</td>
<td>19</td>
</tr>
<tr>
<td>75</td>
<td>6.28</td>
<td>1.97</td>
<td>30</td>
</tr>
<tr>
<td>75</td>
<td>6.05</td>
<td>1.85</td>
<td>27</td>
</tr>
<tr>
<td>75</td>
<td>6.11</td>
<td>1.82</td>
<td>28</td>
</tr>
<tr>
<td>100</td>
<td>6.79</td>
<td>1.93</td>
<td>34</td>
</tr>
<tr>
<td>100</td>
<td>6.68</td>
<td>1.98</td>
<td>32</td>
</tr>
<tr>
<td>100</td>
<td>6.85</td>
<td>1.96</td>
<td>33</td>
</tr>
<tr>
<td>125</td>
<td>6.95</td>
<td>2.10</td>
<td>38</td>
</tr>
<tr>
<td>125</td>
<td>7.10</td>
<td>2.04</td>
<td>38</td>
</tr>
<tr>
<td>125</td>
<td>6.92</td>
<td>2.01</td>
<td>39</td>
</tr>
<tr>
<td>150</td>
<td>7.34</td>
<td>2.23</td>
<td>42</td>
</tr>
<tr>
<td>150</td>
<td>7.41</td>
<td>2.05</td>
<td>41</td>
</tr>
<tr>
<td>150</td>
<td>7.28</td>
<td>2.07</td>
<td>44</td>
</tr>
<tr>
<td>175</td>
<td>7.47</td>
<td>2.11</td>
<td>47</td>
</tr>
<tr>
<td>175</td>
<td>7.45</td>
<td>2.22</td>
<td>48</td>
</tr>
<tr>
<td>175</td>
<td>7.70</td>
<td>2.26</td>
<td>49</td>
</tr>
<tr>
<td>200</td>
<td>7.86</td>
<td>2.38</td>
<td>55</td>
</tr>
<tr>
<td>200</td>
<td>7.93</td>
<td>2.22</td>
<td>54</td>
</tr>
<tr>
<td>200</td>
<td>7.76</td>
<td>2.16</td>
<td>49</td>
</tr>
<tr>
<td>340</td>
<td>9.38</td>
<td>2.47</td>
<td>98</td>
</tr>
<tr>
<td>490</td>
<td>9.72</td>
<td>2.48</td>
<td>107</td>
</tr>
</tbody>
</table>

However, the first three increments of 5, 10 and 20 cm were put in place so that the impact of the ball at relatively low speeds, can be calculated more accurately. The final two increments of height, 340 cm and 490 cm, were somewhat arbitrary, the height that I could drop a ball while standing on a chair and from the roof, respectively. These results were difficult to obtain, as there were problems with accurately landing the ball into the sand, as well as the large amount of sand spillage that took place. As such, there could only be one trial of each recorded, and these increments would be used as a rough indication of whether the trend continued outside of the recorded data.

Mathematical Process

The data was then averaged across each of the heights to eliminate the random error that caused variation between them, using the following formula:

\[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \]

For example, the average diameter of the crater when the golf ball was dropped from 5cm was worked out as such:

\[ \bar{x} = \frac{3.71 + 3.55 + 3.54}{3} = 3.60 \text{cm} \]

The average scores for each variation of measurement vs. height were tabulated in Table 2.
Table 2: Averaged data for different drop heights

<table>
<thead>
<tr>
<th>Drop Height (cm)</th>
<th>Diameter of Crater (cm)</th>
<th>Depth of Crater (cm)</th>
<th>Sand Displaced (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.60</td>
<td>0.82</td>
<td>4.33</td>
</tr>
<tr>
<td>10</td>
<td>4.17</td>
<td>0.99</td>
<td>7.33</td>
</tr>
<tr>
<td>20</td>
<td>4.72</td>
<td>1.18</td>
<td>11.33</td>
</tr>
<tr>
<td>50</td>
<td>5.41</td>
<td>1.56</td>
<td>19.33</td>
</tr>
<tr>
<td>75</td>
<td>6.15</td>
<td>1.88</td>
<td>28.33</td>
</tr>
<tr>
<td>100</td>
<td>6.77</td>
<td>1.96</td>
<td>33</td>
</tr>
<tr>
<td>125</td>
<td>6.99</td>
<td>2.05</td>
<td>39</td>
</tr>
<tr>
<td>150</td>
<td>7.34</td>
<td>2.12</td>
<td>42.3</td>
</tr>
<tr>
<td>175</td>
<td>7.54</td>
<td>2.20</td>
<td>48</td>
</tr>
<tr>
<td>200</td>
<td>7.85</td>
<td>2.25</td>
<td>52.7</td>
</tr>
<tr>
<td>340</td>
<td>93.8</td>
<td>24.7</td>
<td>98</td>
</tr>
<tr>
<td>460</td>
<td>97.2</td>
<td>24.8</td>
<td>107</td>
</tr>
</tbody>
</table>

However, the results gleaned from the experiment were limited in that they showed only the weight of sand required to refill in the crater rather than volume.

To find the volume of the crater the relative density of the sand used must be calculated using the following formula:

\[ \text{density} = \frac{\text{mass}}{\text{volume}} \]

It was shown that 125 cm³ of sand was found to weigh 196g. Hence the density of sand used was calculated as:

\[ 196 \div 125 = 1.568 \text{g.cm}^{-3} \]

Therefore, by rearranging the formula, the volume could be deduced from weight according to the formula:

\[ \text{volume} = \frac{\text{mass}}{\text{density}} \]

For example, the average crater volume made by a 5cm drop height was calculated thus:

\[ 4.33 \div 1.568 = 2.761 \text{g.cm}^{-3} \]

The crater volumes for each drop height were therefore calculated in Table 3.

Table 3: Height vs Calculated crater volume

<table>
<thead>
<tr>
<th>Drop Height (cm)</th>
<th>Average Crater Volume (cm³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2.764</td>
</tr>
<tr>
<td>10</td>
<td>4.677</td>
</tr>
<tr>
<td>20</td>
<td>7.288</td>
</tr>
<tr>
<td>50</td>
<td>12.330</td>
</tr>
<tr>
<td>75</td>
<td>18.070</td>
</tr>
<tr>
<td>100</td>
<td>21.046</td>
</tr>
<tr>
<td>125</td>
<td>24.872</td>
</tr>
<tr>
<td>150</td>
<td>26.998</td>
</tr>
<tr>
<td>175</td>
<td>30.612</td>
</tr>
<tr>
<td>200</td>
<td>33.588</td>
</tr>
<tr>
<td>340</td>
<td>62.500</td>
</tr>
<tr>
<td>490</td>
<td>68.240</td>
</tr>
</tbody>
</table>

The second problem to be solved is the conversion from drop height to speed at impact. This is converted through the combination of two formulas. The first calculates the length of time the ball is falling from a certain height and the second will give the speed of the ball at a known time (as calculated by the first formula):

Formula 1:

\[ h = V_i + \frac{1}{2} gt^2 \]
Where

\[ h = \text{Height} \]
\[ V_i = \text{Initial velocity} \]
\[ g = \text{Gravity} \]
\[ t = \text{Time} \]

Since the ball was at rest before being dropped, its initial velocity \( V_i \) is considered as 0. We also know that gravity is \( 9.81 \text{m/s}^2 \).

\[ t = \sqrt{\frac{2h}{9.81 \times 100}} \]

However, we must also note that we must convert cm to m and therefore divide by 100. The first formula is therefore rearranged as follows:

\[ t = \sqrt{\frac{h}{9.81 \times 50}} \]

From the time spent in the air, we can then determine the speed of the ball through gravity by \( 9.81 \text{m/s}^2 \). This is given by the following formula:

Formula 2:
\[ V_f = V_i + g \times t \]

Where

\( V_f = \text{Final velocity} \)
\( V_i = \text{Initial velocity} \)
\( g = \text{Gravity} \)
\( t = \text{Time} \)

Once again, \( V_i = 0 \) and hence in this experiment \( V_f = g \times t \). Combining the two formulas we get:

\[ V_f = g \times \sqrt{\frac{h}{9.81 \times 50}} \]

For example the ball dropped at 5cm will have the following velocity:

\[ V_f = 9.8 \times \sqrt{\frac{5}{9.81 \times 50}} = 0 \]

This can be tabulated as seen in Table 4:

<table>
<thead>
<tr>
<th>Drop Height (cm)</th>
<th>Drop Speed (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.99</td>
</tr>
<tr>
<td>10</td>
<td>1.40</td>
</tr>
<tr>
<td>20</td>
<td>1.98</td>
</tr>
<tr>
<td>50</td>
<td>3.13</td>
</tr>
<tr>
<td>75</td>
<td>3.83</td>
</tr>
<tr>
<td>100</td>
<td>4.43</td>
</tr>
<tr>
<td>125</td>
<td>4.95</td>
</tr>
<tr>
<td>150</td>
<td>5.42</td>
</tr>
<tr>
<td>175</td>
<td>5.86</td>
</tr>
<tr>
<td>200</td>
<td>6.26</td>
</tr>
<tr>
<td>340</td>
<td>8.16</td>
</tr>
<tr>
<td>490</td>
<td>9.80</td>
</tr>
</tbody>
</table>
The Correlation between Drop Height and Crater Size

Scatter plots:
Scatter plot diagrams allow me to see if there is a trend in results; it will also allow the plotting of a line of best fit if appropriate.

Graph 1: Drop vs Depth

Visually, there seems to be a strong positive correlation between all of the different variables, as expected. The diameter vs. depth graph shows a linear relationship with a correlation between the depth and diameter of a crater and hence a consistent shape to the craters. The other graphs have a non-linear correlation, which indicates the velocity of the golf ball would be related to the square of its height above the ground as described in Table 4. When speed rather than drop height is used to compare to the crater diameter, depth and volume, the following graphs are achieved:

Graph 2: Drop vs Diameter

Graph 3: Diameter vs Depth

Graph 4: Height vs Volume

Graph 5: Speed vs Diameter
Graph 6: Speed vs Depth

Graph 7: Speed vs Volume

Visually, these graphs seem to be much more linear in shape than the previous ones, which makes sense as the ball accelerates towards the ground, thus having diminishing time in the air and therefore less time to speed up.

Finding the Correlation Coefficient

As my intent is to find the strength of the relationship between the speed of the ball and the other variables of a crater’s size, I will find the correlation coefficients using the Pearson’s correlation test. These will give an indication about whether there is indeed a strong linear relationship between the two factors. If there indeed is, I will find a ‘line of best fit’ from which I can predict results. The formula below is used to calculate the Coefficient of correlation:

\[
r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 (y - \bar{y})^2}}
\]

Using the data above for the speed vs crater volume, I have generated the following table 5.

Table 5. Speed vs Volume

<table>
<thead>
<tr>
<th></th>
<th>x - (\bar{x})</th>
<th>y - (\bar{y})</th>
<th>(x - (\bar{x}))(y - (\bar{y}))</th>
<th>(x - (\bar{x}))^2</th>
<th>(y - (\bar{y}))^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.695</td>
<td>-23.313</td>
<td>86.132</td>
<td>13.650</td>
<td>543.493</td>
<td></td>
</tr>
<tr>
<td>-2.705</td>
<td>-18.850</td>
<td>50.981</td>
<td>7.315</td>
<td>355.306</td>
<td></td>
</tr>
<tr>
<td>-1.554</td>
<td>-13.748</td>
<td>21.364</td>
<td>2.415</td>
<td>188.995</td>
<td></td>
</tr>
<tr>
<td>-0.850</td>
<td>-8.008</td>
<td>6.810</td>
<td>0.723</td>
<td>64.121</td>
<td></td>
</tr>
<tr>
<td>-0.257</td>
<td>-5.032</td>
<td>1.295</td>
<td>0.066</td>
<td>25.316</td>
<td></td>
</tr>
<tr>
<td>0.265</td>
<td>-1.205</td>
<td>-0.319</td>
<td>0.070</td>
<td>1.451</td>
<td></td>
</tr>
<tr>
<td>0.738</td>
<td>0.921</td>
<td>0.680</td>
<td>0.544</td>
<td>0.849</td>
<td></td>
</tr>
<tr>
<td>1.172</td>
<td>4.535</td>
<td>5.316</td>
<td>1.374</td>
<td>20.570</td>
<td></td>
</tr>
<tr>
<td>1.576</td>
<td>7.511</td>
<td>11.841</td>
<td>2.485</td>
<td>56.422</td>
<td></td>
</tr>
<tr>
<td>3.479</td>
<td>36.423</td>
<td>126.710</td>
<td>12.102</td>
<td>1326.668</td>
<td></td>
</tr>
<tr>
<td>5.115</td>
<td>42.162</td>
<td>215.680</td>
<td>26.168</td>
<td>1777.672</td>
<td></td>
</tr>
<tr>
<td>SUM:</td>
<td>596.781</td>
<td>77.701</td>
<td>4818.847</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hence:

\[
r = \frac{596.781}{\sqrt{77.701 \times 4818.847}} = 0.9753
\]

The value of \(r = 0.9753\) means that there is indeed a very strong, linear, positive relationship between the ball’s velocity at impact and the size of the crater it made. Due to the strength of the correlation, I tried to find the line of regression to predict future results as well as the \(r^2\) value to test
the degree to which speed “explains” volume in craters. After putting it into Microsoft Excel, the $r^2$ value was determined to be:

$$r^2 = (0.9753)^2 = 0.9512$$

The $r^2$ value, shows that 95.12% of the crater’s volume can be explained by the speed at which the object impacts the ground, all other things remaining the same, and means that it is a highly relevant factor in the size, in volume, of the resultant crater.

**Comparison to Ideal Shapes**

To complete my secondary mission of deducing what 3D model best describes the shape of the crater, I will first have to calculate the volumes of ideal shapes based on their circumference and depths. For continuity, each of the examples shown will be the average depths and diameters of the drop height of 5 cm, those being:

Diameter = 3.6 cm and depth = 0.82 cm

The formula for the volume of an **elliptical paraboloid** is:

$$V = \frac{1}{2} \pi (r)^2 h$$

Where

$V$ = Volume  
$r$ = Radius  
$h$ = Height/Depth

The formula for the volume of a **cone** is:

$$V = \frac{1}{3} \pi (r)^2 h$$

Where

$r$ = Radius  
$h$ = Depth

$$V = \frac{1}{3} \pi (1.8)^2 \times 0.82 = 2.782 cm^3$$

The formula for the volume of the **cap of a sphere** is:

$$V = \frac{1}{6} \pi r (3a^2 + h^2)$$

Where

$a$ = Radius  
$h$ = Height/Depth

$$V = \frac{1}{6} \pi \times 0.82 (3 \times 1.8^2 + 0.82^2) = 4.462 cm^3$$

Using these formulae, the following table was created comparing the predicted volumes for each crater shape to the measured (actual volumes). Therefore the results, when tabulated, look like this in Table 6.
Table 6: Comparing different shape volumes

<table>
<thead>
<tr>
<th>Crater Diameter (cm)</th>
<th>Sphere Volume (cm³)</th>
<th>Cone Volume (cm³)</th>
<th>Parabola Volume (cm³)</th>
<th>Measured Crater Volume (cm³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6</td>
<td>4.46</td>
<td>2.78</td>
<td>4.17</td>
<td>2.76</td>
</tr>
<tr>
<td>4.17</td>
<td>7.28</td>
<td>4.51</td>
<td>6.77</td>
<td>4.68</td>
</tr>
<tr>
<td>4.72</td>
<td>11.18</td>
<td>6.88</td>
<td>10.32</td>
<td>7.23</td>
</tr>
<tr>
<td>5.41</td>
<td>19.95</td>
<td>11.96</td>
<td>17.95</td>
<td>12.33</td>
</tr>
<tr>
<td>6.15</td>
<td>31.37</td>
<td>18.60</td>
<td>27.89</td>
<td>18.07</td>
</tr>
<tr>
<td>6.77</td>
<td>39.17</td>
<td>23.50</td>
<td>35.25</td>
<td>21.05</td>
</tr>
<tr>
<td>6.99</td>
<td>43.84</td>
<td>26.22</td>
<td>39.33</td>
<td>24.87</td>
</tr>
<tr>
<td>7.34</td>
<td>49.79</td>
<td>29.88</td>
<td>44.82</td>
<td>27.00</td>
</tr>
<tr>
<td>7.54</td>
<td>54.59</td>
<td>32.69</td>
<td>49.04</td>
<td>30.61</td>
</tr>
<tr>
<td>7.85</td>
<td>60.52</td>
<td>36.35</td>
<td>54.53</td>
<td>33.59</td>
</tr>
<tr>
<td>9.38</td>
<td>93.23</td>
<td>56.89</td>
<td>85.34</td>
<td>62.50</td>
</tr>
<tr>
<td>9.72</td>
<td>100.00</td>
<td>61.34</td>
<td>92.01</td>
<td>68.24</td>
</tr>
</tbody>
</table>

These results were represented graphically against the actual volume of the crater as calculated by the weight of the sand (multiplied by the density) required to fill it back in.

Visually, it looks as if the volume of the recorded craters most closely fits the predictions made by the shape of a cone. To prove this similarity mathematically, I made a comparison of the averages of each measurement, comparing the average percentage error.

Percentage error is the most appropriate calculation for this, as it does not take into account the literal size of the error, but the error’s relative size. It can be found using the following formula.

$$\frac{\text{estimated} - \text{real}}{\text{real}}$$

However, since it does not matter whether the estimated volume, that of the ideal shape, is larger or smaller than the ‘real’ volume of the craters, an absolute value can be made to keep all numbers positive. This means altering the formula so it looks like this.

$$\sqrt{\frac{(\text{estimated} - \text{real})^2}{\text{real}}}$$

**Comparison of Ideal Shapes vs Measured Results**

**Graph 9: Comparing Volumes Graphically**

There is obviously a far lower percentage error on the shape of the cone (6%) than any other ideal shape shown. Thus, it should be accepted that the shape of the craters made were best described by cones.
Interpretation of Results
To investigate the relationship between the speed of a falling object and the shape of the resultant crater it makes when it impacts the ground, I gathered data from a rig that I made. I measured the diameter and depth of the crater made by an object landing in dry sand as well as the weight of sand displaced by the impact. I then found the volume of each crater using the known density of the sand. After averaging the results, I then determined what speed the ball was travelling at its point of impact. To determine the nature of the relationship between the two variables I used the Pearson’s correlation co-efficient. My calculations show that the Pearson’s Correlation coefficient equalled 0.9753, which corresponded to a very strong, positive, linear correlation, thus proving there was a strong correlation between the speed of a falling object and crater dimension it creates. This is somewhat of an obvious result and is due to the fact that an object picks up speed, and thus momentum as it falls, making the force with which it hits the ground greater as the height increases. The value if this result, however, is that the linear relationship can be used either to predict any future crater based on the drop height of an object or, conversely, that velocity of a falling object can be deduced by the crater it had created. I then attempted to complete my secondary aim of determining the shapes of the craters I had made. To do this, I first calculated the volumes of 3 ideal shapes, namely an elliptical paraboloid, a sphere cap and a cone, using the diameter and depth of the average craters that I had measured. I then compared the volumes of these ideal shapes against the volumes that were measured using the rig I had set up. The comparison was done by calculating the average percentage difference between the measured volume and each of the ideal shapes. The result showed, both visually and mathematically, that the shape of the craters were best described as a cone. This was surprising to me as I had expected, looking at the photos of moon and earth impact craters, that the crater would be a paraboloid, if not a perfect sphere cap. This may be due to the fact that, while the very end of the crater was curved, where the ball actually impacted, there was no reason that the rest of the crater should have been curved like a sphere, as the rounded ball never touched it. This can be seen by the fact that the volumes of the paraboloid and sphere cap diverged from the cone as the crater grew larger and thus the ball itself did not touch as much of it.

Table 7: Resultant data

<table>
<thead>
<tr>
<th>Drop Height (cm)</th>
<th>Cap % Error</th>
<th>Paraboloid % Error</th>
<th>Cone % Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>61.5</td>
<td>51.0</td>
<td>0.7</td>
</tr>
<tr>
<td>10.0</td>
<td>55.6</td>
<td>44.8</td>
<td>3.5</td>
</tr>
<tr>
<td>20.0</td>
<td>54.7</td>
<td>42.8</td>
<td>4.8</td>
</tr>
<tr>
<td>50.0</td>
<td>61.8</td>
<td>45.5</td>
<td>3.0</td>
</tr>
<tr>
<td>75.0</td>
<td>73.6</td>
<td>54.4</td>
<td>2.9</td>
</tr>
<tr>
<td>100.0</td>
<td>86.1</td>
<td>67.5</td>
<td>11.7</td>
</tr>
<tr>
<td>125.0</td>
<td>76.3</td>
<td>58.1</td>
<td>5.4</td>
</tr>
<tr>
<td>150.0</td>
<td>84.4</td>
<td>66.0</td>
<td>10.7</td>
</tr>
<tr>
<td>175.0</td>
<td>78.3</td>
<td>6.2</td>
<td>6.8</td>
</tr>
<tr>
<td>200.0</td>
<td>80.2</td>
<td>62.3</td>
<td>8.2</td>
</tr>
<tr>
<td>340.0</td>
<td>49.2</td>
<td>36.5</td>
<td>9.0</td>
</tr>
<tr>
<td>490.0</td>
<td>46.5</td>
<td>34.8</td>
<td>10.1</td>
</tr>
<tr>
<td>4400.0</td>
<td>73.8</td>
<td>59.3</td>
<td>6.2</td>
</tr>
<tr>
<td>4700.0</td>
<td>61.3</td>
<td>48.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Average % Error</td>
<td>67.4</td>
<td>52.3</td>
<td>6.0</td>
</tr>
</tbody>
</table>
Validation
While I feel that, overall, the experiment and mathematics were a great success in determining the shape of a crater, its validity was limited. Firstly, while it is true that light silicates i.e. sand make up most of the earth's surface, the filtered sand that was used for the experiment was only a rough and extremely simplified model of the earth's surface. On top of this, the sand had a tendency to slide down the side of the crater when the golf ball was removed, thus leaving the possibility that it changed the shape of the crater. Experimentation with other light and granulated materials could provide the key to eliminating such possibility of error. The second area of possible improvement could have been the accuracy of measurements. While the electrical Vernier callipers and electrical scales were themselves precise, problems arose, especially with gauging the depth of the crater and its volume. While precision was obviously strived for, it is difficult to know the extent of uncertainty with which the measurements could be taken, especially when measuring the fractions of millimetres that the smaller craters required to differentiate themselves. However, without much more expensive equipment, it would be relatively impossible to have achieved a more accurate outcome. Finally, I was limited to a height of 4m for a drop height. The resultant graphs may show a straight line correlation at the test speeds but may show a less straight line correlation if impact speeds are much higher. The meteor causing the extinction of the dinosaurs, for example, would have been travelling at many kilometres per second rather than merely several meters per second.

Step 1: Drop ball from fixed height into sand

Step 2: Remove ball.

Step 3: Scrape off excess sand and measure diameter of crater with callipers.

Step 4: Measure depth of crater.
There cannot be an Odd Perfect Number using the Properties of Odds and Evens

Flynn Innes
Trinity Grammar School, Summer Hill NSW, Australia
E-mail: finnes.2018@trinity.nsw.edu.au

Abstract
Mathematical inquiry can often lead to a jungle of unique questions and problems. In the field of Number Theory, there are a wide assortment of such mathematical creatures. Although these problems are easy to state, they can remain dormant for years with little sign of progress. In this article, we show that there cannot be an Odd Perfect Number using the Properties of Odds and Evens.

A perfect numbers are positive integers that are the sum of their proper divisors. For instance, 6 is a perfect number, because the sum of its proper divisors, 1, 2, and 3 equals 6 \((1 + 2 + 3 = 6)\). A perfect number may be expressed as the sum of all its factors (excluding the number itself); for example, the number 6 may be expressed as \(1 + 2 + 3\).

Look at the following cases:

Odd * Odd = Odd
Odd * Even = Even
Even*Even=Even

Since Odd * Odd is the only case which results in an Odd, all Odd numbers can only have Odd factors.

Next we must prove that there are no perfect numbers, which are perfect squares.

Take a perfect number \(n\) and count how many odd factors it has; let \(k\) be the number of odd factors, \(p\) be the prime factor and \(e\) be the power which to raise \(p\) to. Then the prime factorization is

\[
n = 2^{e_0} p_1^{e_1} p_2^{e_2} \ldots p_k^{e_k}
\]

Therefore, the number of odd divisors of \(n\) is \((e_1 + 1)(e_2 + 1)\ldots(e_k + 1)\)

If \(n\) is a perfect square, then each of the exponents in the prime factorization of \(n\) must be even. Thus, with an even number of odd factors by the laws of odds and evens the number of odd divisors of \(n\) must be odd.

If \(n\) is even, then the sum of all the divisors must be an odd number, which by definition contradicts the assumption that \(n\) is perfect.

If \(n\) is odd, then \(n\) must have an even number of odd proper divisors, thus the sum of all of the proper divisors of \(n\) is an even number, which means by definition \(n\) cannot be perfect.

Therefore, if \(n\) is a perfect number, \(n\) cannot be a perfect square.

Since a perfect number cannot be a perfect square as proven above, all perfect numbers must have an even number of factors.

Since there is an even number of odd factors, the sum of all the factors MUST be an even number.

Since the sum of the factors MUST be even and the perfect number itself is odd, by definition it cannot be a perfect number.

Therefore there are no odd perfect numbers.