The Mathematics of Sudoku

Turner King

Abstract: We find the least possible number of squares that can be removed from a completed Sudoku puzzle to yield more than one possible solution.

Introduction

Sudoku is latest puzzle craze to hit the world. Despite its Japanese name, it was originally released as in the United States in 1979. However, the puzzle first caught fire when it was released in Japan in 1984. More recently, it has garnered international success with its simple concept put into a complex puzzle.

Board & Rules

A traditional Sudoku puzzle is a 9x9 square grid separated into nine 3x3 regions. However, there are versions of Sudoku played with larger or smaller grids and with and without regions. Given several numbers put in to start, the goal of the puzzles is to insert numbers (1-9 in the case of the 9x9 grid), into the squares such that each number only appears once in each row, in each column, and in each region (3x3 in this case).

Methods to Solve Objective

To solve for the least possible number of squares that need to be removed from a Sudoku puzzle to create more than one solution, I had to create my own complete puzzles, and by a trial and error method, give an answer to my question. I first had to figure out how to make complete puzzles. The method of plugging in numbers in spots and trying fix up the puzzle later proved to be simply frustrating and time consuming, but before too long I constructed my first puzzle using the variables *a-j* for the numbers 1-9.

a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*
d	*e*	*f*	*g*	*h*	*i*	*a*	*b*	*c*
g	*h*	*i*	*a*	*b*	*c*	*d*	*e*	*f*
b	*c*	*a*	*e*	*f*	*d*	*h*	*i*	*g*
e	*f*	*d*	*h*	*i*	*g*	*b*	*c*	*a*
h	*i*	*g*	*b*	*c*	*a*	*e*	*f*	*d*
c	*a*	*b*	*f*	*d*	*e*	*i*	*g*	*h*
f	*d*	*e*	*i*	*g*	*h*	*c*	*a*	*b*
i	*g*	*h*	*c*	*a*	*b*	*f*	*d*	*e*

The algorithm that I used to create the first puzzle was to take the variables *a-j* and put them into the first row. Then, I separated them into groups of three variables *a-c, d-f,* and *g-i* respectively. For the next two groups, I shifted them left one region creating a unique row. At the start of next set of regions, within each group of three I shifted variables one to the left within their groups and proceeded to finish the puzzle by doing these rules.

Original Possibility

When I completed the first puzzle that I realized my first possible solution to the problem, I found that within the puzzle if you exchange all nine of one variable for the nine of another you still end up with a legal puzzle.

a	b	c	d	e	f	g	**h**	**i**
d	e	f	g	**h**	**i**	a	b	c
g	**h**	**i**	a	b	c	d	e	f
b	c	a	e	f	d	**h**	**i**	g
e	f	d	**h**	**i**	g	b	c	a
h	**i**	g	b	c	a	e	f	d
c	a	b	f	d	e	**i**	g	**h**
f	d	e	**i**	g	**h**	c	a	b
i	g	**h**	c	a	b	f	d	e

The variables h & i are interchanged.

a	b	c	d	e	f	g	**i**	**h**
d	e	f	g	**i**	**h**	a	b	c
g	**i**	**h**	a	b	c	d	e	f
b	c	a	e	f	d	**i**	**h**	g
e	f	d	**i**	**h**	g	b	c	a
i	**h**	g	b	c	a	e	f	d
c	a	b	f	d	e	**h**	g	**i**
f	d	e	**h**	g	**i**	c	a	b
h	g	**i**	c	a	b	f	d	e

So if rather, these two variables are completely deleted rather than switched you end up with a puzzle with more than one solution, two to be exact. The exact same thing happens with two rows or columns within the same set of regions.

a	b	c	d	e	f	g	h	i
d	e	f	g	h	i	a	b	c
g	h	i	a	b	c	d	e	f
b	c	a	e	f	d	h	i	g
e	f	d	h	i	g	b	c	a
h	i	g	b	c	a	e	f	d
c	a	b	f	d	e	i	g	h
f	**d**	**e**	**i**	**g**	**h**	**c**	**a**	**b**
i	**g**	**h**	**c**	**a**	**b**	**f**	**d**	**e**

The bottom two rows are switched.

a	b	c	d	e	f	g	h	i
d	e	f	g	h	i	a	b	c
g	h	i	a	b	c	d	e	f
b	c	a	e	f	d	h	i	g
e	f	d	h	i	g	b	c	a
h	i	g	b	c	a	e	f	d
c	a	b	f	d	e	i	g	h
i	**g**	**h**	**c**	**a**	**b**	**f**	**d**	**e**
f	**d**	**e**	**i**	**g**	**h**	**c**	**a**	**b**

Both of these yield a solution of 18 for the number squares that need to be deleted to result in multiple solutions for the puzzle, which is good, but as we'll find out is a lot more than the actual solution.

Actual Solution

After pondering the question the further, I got the idea of pairing variables as such so I would only need four squares deleted to create two unique solutions. After several attempts at creating a legal puzzle, I finally found a solution.

a	*d*	*f*	*b*	*h*	*c*	*g*	*e*	***i***
i	*h*	*c*	*d*	*e*	*g*	*f*	*b*	***a***
b	*e*	*g*	***a***	*f*	***i***	*h*	*d*	*c*
f	*a*	*d*	*c*	*b*	*h*	*i*	*g*	*e*
c	*i*	*h*	*g*	*d*	*e*	*a*	*f*	*b*
g	*b*	*e*	*i*	*a*	*f*	*c*	*h*	*d*
d	*f*	*a*	*h*	*c*	*b*	*e*	*i*	*g*
h	*c*	*i*	*e*	*g*	*d*	*b*	*a*	*f*
e	*g*	*b*	*f*	*i*	*a*	*d*	*c*	*h*

If the *a* and *i* are deleted from the top left and right corners we are left with an impossible puzzle to solve since the remaining bolded *a* and *i* are in the remaining row of the three in the set of three regions and cannot therefore induce any of the *a's* and *i's* we deleted from the puzzle. An interesting correlation that I found with deleting the appropriate squares is the fact that we are deleting two squares from two rows, two columns, and two regions. I believe that it is this correlation that makes this possible. Now since whenever two squares are in the same row with each other they cannot be in the same column and vice-versa. I'll use this to prove that no smaller case is possible. For example, in the case of deleting three squares, if we put all three in the same row, they are all induced by each of their columns still creating a unique solution. In the case where there two in one row and one in another, that one is induced by its column, then the remaining two are induced by their columns. Finally in the case where they're all in different rows, their rows will induce the remaining needed variable. With two squares, if they are placed in the same row, both squares can be induced by their columns, and with them in separate rows, their rows will induce their solution. Lastly, in the obvious case of removing one square that one square can be induced by the row, column, or region, it's in.

Proof

In a 9x9 grid with nine 3x3 regions, a Sudoku puzzle can be produced, so that no solution can be found by simply removing four squares.

Further Research

I've pondered a lot of questions when working with these Sudoku puzzles. One question that I came up with is in ways the opposite of the one I presented here. How many squares do you need to place initially to ensure that you have only one solution possible? Another question is if the solution to this research works with other sizes of grids. One last question I'd like to present is solution for the total number of puzzles that can be produced. I believe that this last one would a very interesting and complex counting problem.