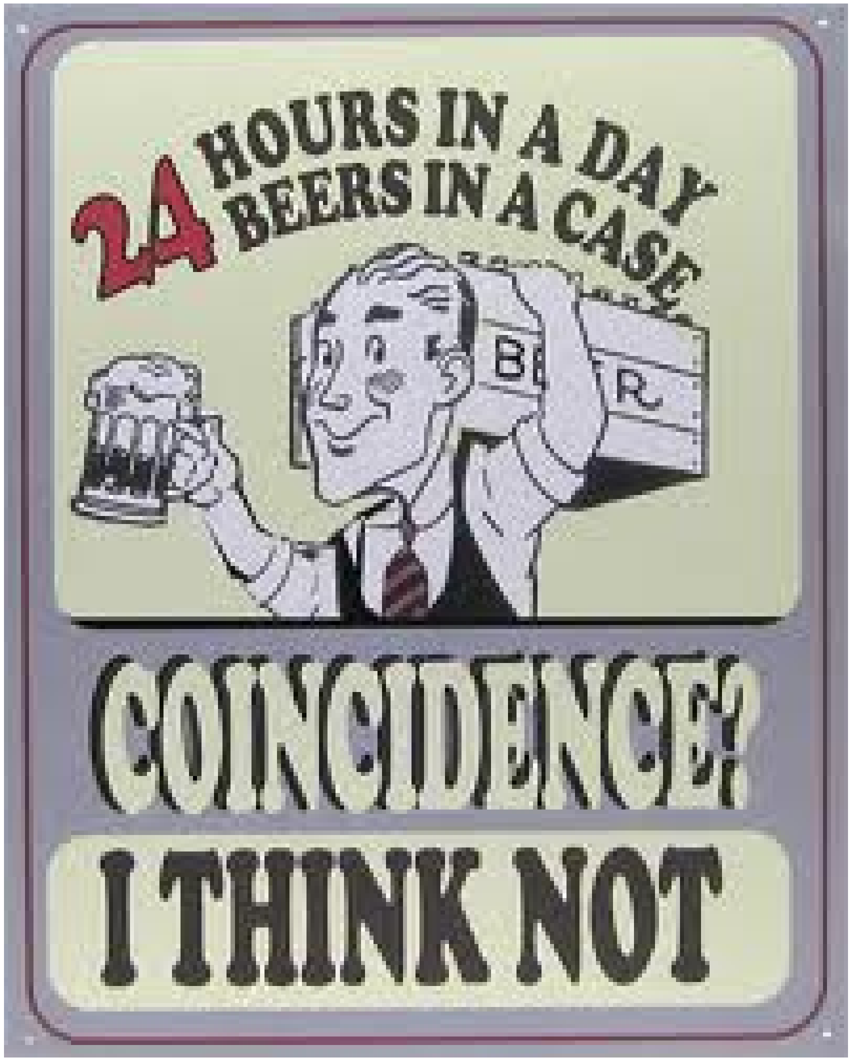
Coincidence in Runs and Clusters

MILO SCHIELD
Augsburg College
Director, W. M. Keck Statistical Literacy Project
US Rep, International Statistical Literacy Project
Member, International Statistical Institute
President, ASA Twin Cities Chapter

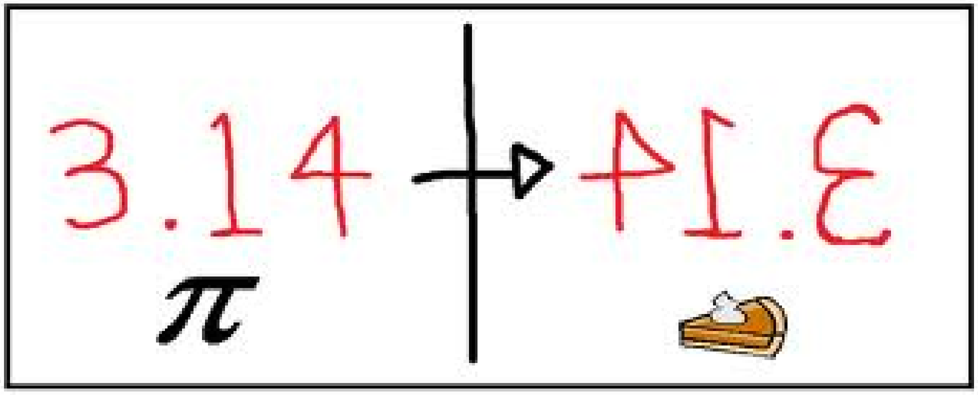
March 8, 2012

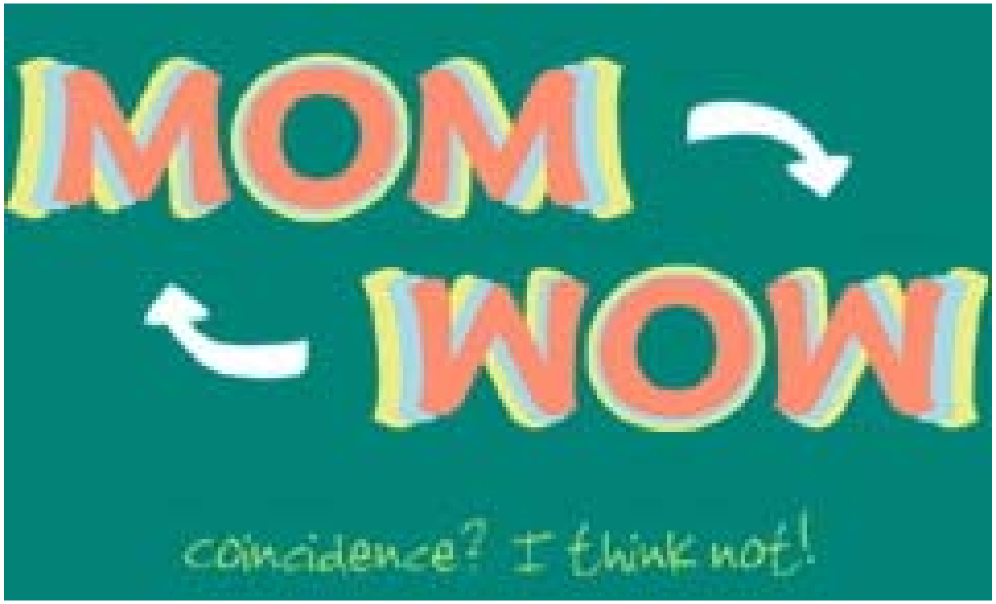
Paper at www.StatLit.org/pdf/2012Schield-MAA.pdf
Slides at www.StatLit.org/pdf/2012Schield-StatChat6up.pdf

Coincidence?



Coincidence?





March 2012

Chance: run of 10 heads?
One chance in 2^10 = 1 in 1,024

Three Heads in a Row

* has one chance in eight. P = 1/8

* is *expected* in 8 sets of three. N = 1/(1/8) = 8.

Binomial distribution: N*P = <Expected> = Mean.

If N = 1/P, then <Exp> = 1.

1	0	1		1	0	0
0	1	1		0	0	1
1	1	1		0	1	1
1	1	0		1	1	0

Map 8 sets of 3 each onto 10 tries

Run of 3 heads is generally found in (1/p)^k+(k-1) flips of a fair coin.										Schield (2012) V1	
Coin	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	
RUN											
3	1	0	1	1	1	0	0	1	1	0	
1	1	0	1								Set #1
2		0	1	1							Set #2
3			1	1	1						Set #3
2				1	1	0					Set #4
1					1	0	0				Set #5
1						0	0	1			Set #6
2							0	1	1		Set #7
2								1	1	0	Set #8

Distribution of longest run of heads in a set of 3

Longest Run	0	1	2	3
Expect #	1	4	2	1
Pctg of 8	12.5%	50.0%	25.0%	12.5%
	TTT	HTT, THT, TTH, HTH.	HHT, THH	HHH

Distribution of Longest Run of Successes

N=2: Counts: 1, 2, 1: TT; HT, TH; HH

N = 3: Counts: 1, 4, 2, 1.

0H: TTT. 1H: HTT, THT, TTH, HTH.

2H: HHT, THH. 3H: HHH.

N = 10: 1, 143, 360, 269, 139, 64, 28, 12, 5, 2, 1.

Mode=2, Median=3, Mean = 2.80.

Mean <> 3, but close enough as a rule-of-thumb.

Summary Statistics: Distribution of Longest Runs

K	N	Mode	Median	Mean
3	8	2	2	2.51
4	16	3	3	3.43
5	32	4	4	4.38
6	64	5	5	5.35
7	128	6	6	6.34
8	256	7	7	7.32
9	512	8	8	8.3
10	1,023	9	9	9.26

Runs: Adjacent Events

Law of Very-Large Numbers (Qualitative):
The very unlikely is almost certain given enough tries

RUNS RULE-OF-THUMB:

A run of events with 1 chance in N
is *generally found* in N tries.

90 tries/row, 15 rows: 1450 tries
Run of 10 (or more) is generally found

Conclusion

Students need to "see" that coincidences

1. **are more common than expected**
2. **depend on the context**
3. **compare ex-ante with ex-post**
4. **may still be signs of causation (Cholera)**

That runs with 1 chance in N are generally found in N tries.

Patterns in Rice

With rice scattered in two dimensions, people can often see shapes that are very unlikely.

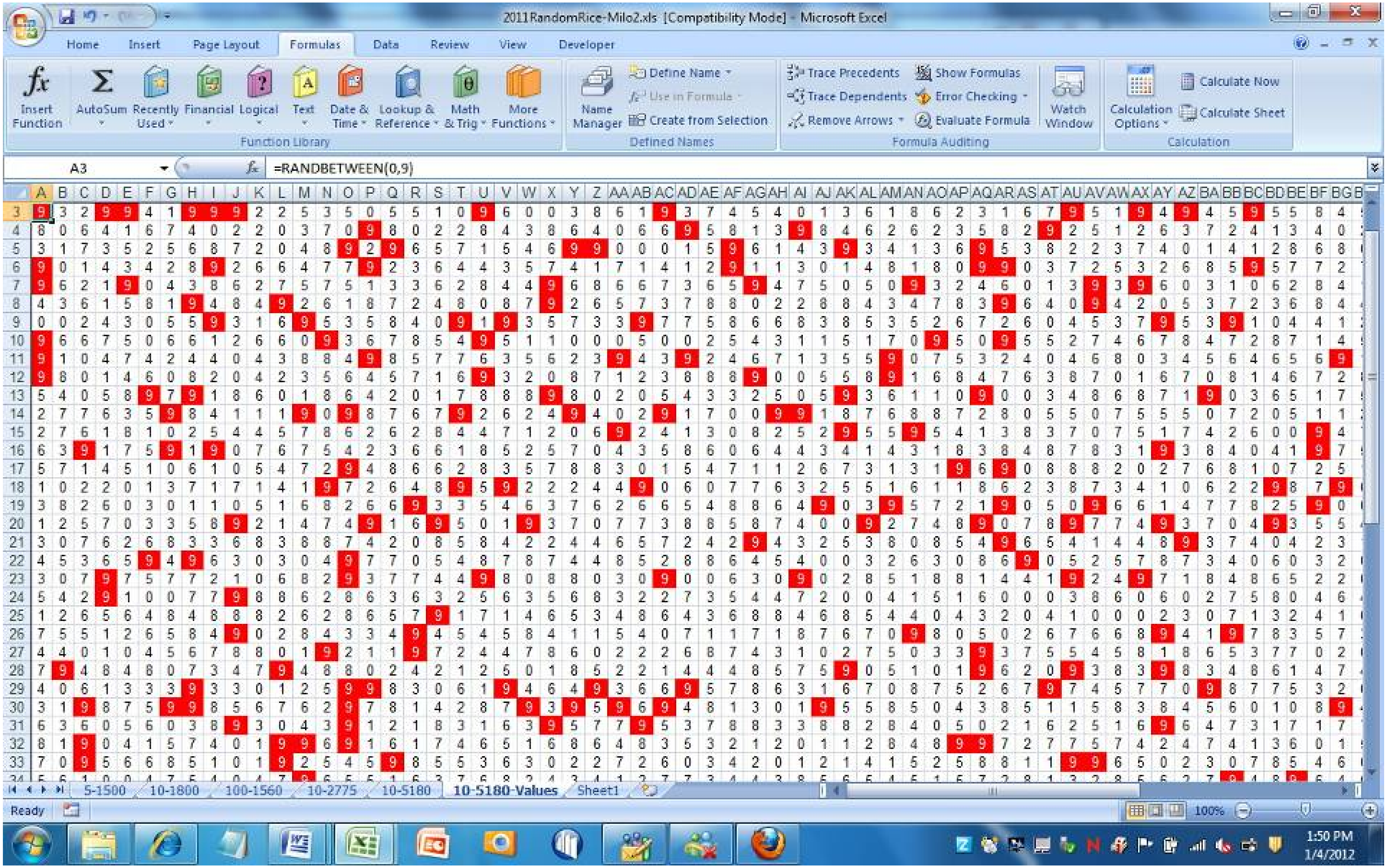
Let’s simulate rice in Excel where each cell has 1 chance in 10.

Patterns in Rice: # Touching
2:1/100; 4:1/10,000; 6: 1/1,000,000

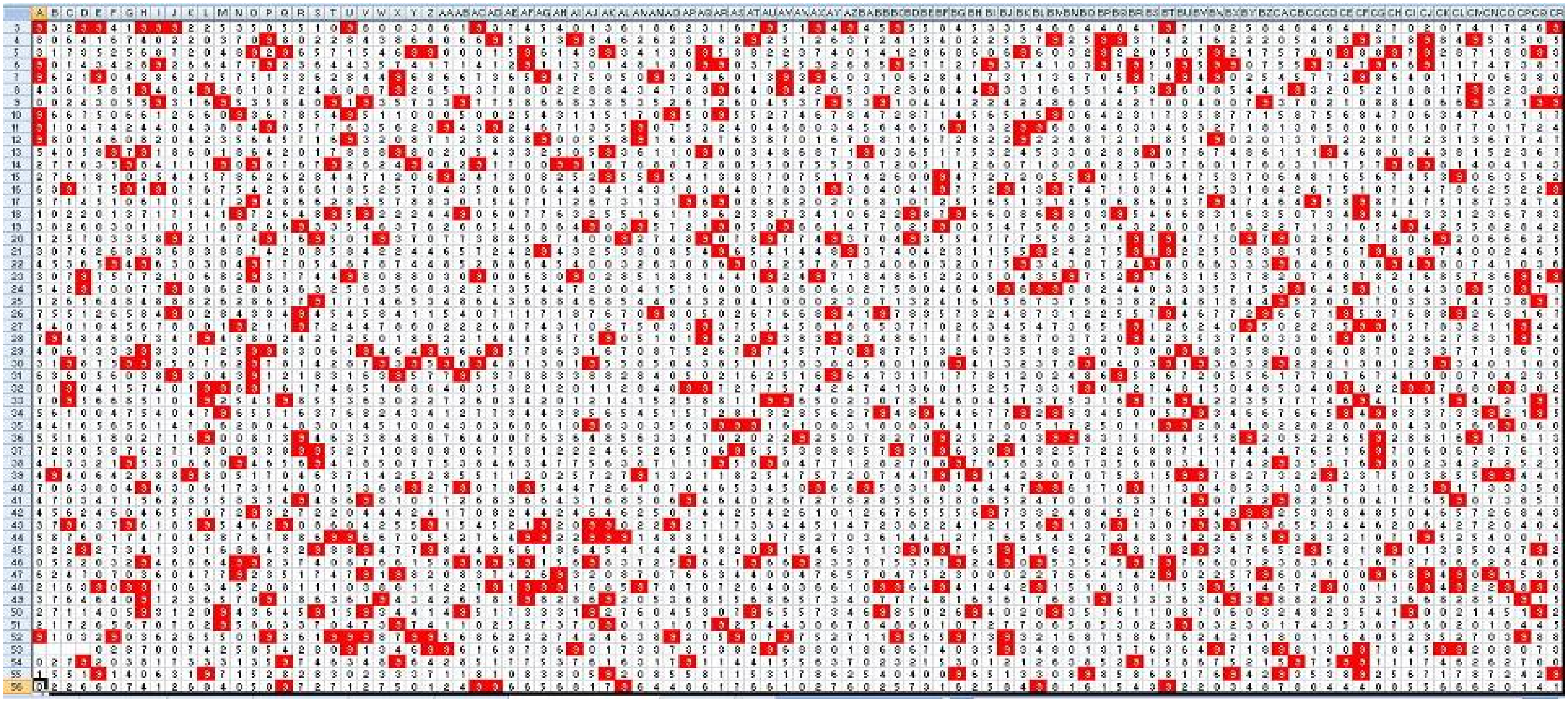
A3 ▼ *fx* =RANDBETWEEN(0,9)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
3	9	3	2	9	9	4	1	9	9	9	2	2	5	3	5	0	5	5
4	8	0	6	4	1	6	7	4	0	2	2	0	3	7	0	9	8	0
5	3	1	7	3	5	2	5	6	8	7	2	0	4	8	9	2	9	6
6	9	0	1	4	3	4	2	8	9	2	6	6	4	7	7	9	2	3
7	9	6	2	1	9	0	4	3	8	6	2	7	5	7	5	1	3	3
8	4	3	6	1	5	8	1	9	4	8	4	9	2	6	1	8	7	2
9	0	0	2	4	3	0	5	5	9	3	1	6	9	5	3	5	8	4
10	9	6	6	7	5	0	6	6	1	2	6	6	0	9	3	6	7	8
11	9	1	0	4	7	4	2	4	4	0	4	3	8	8	4	9	8	5
12	9	8	0	1	4	6	0	8	2	0	4	2	3	5	6	4	5	7

Patterns in Rice



3 touching: 1in 1,000
6 touching: 1 in a million



Patterns in Rice

In 2D, there are more ways for cells to connect:

2 horizontally (left side or right side)

2 vertically (above and below)

4 vertices (NE, SE, SW and NW corners)

8 TOTAL ways two random cells can connect.

Chance that 6 cells with rice will touch:

a.1 in 10^6: 1 in a million

b. (8-1)^6 = 262,144

Conclusion

Students need to "see" that coincidences

1. **are more common than expected**
2. **depend on the context**
3. **compare ex-ante with ex-post**
4. **may still be signs of causation (Cholera)**

That runs with 1 chance in N are generally found in N tries.

The "Birthday" Problem

Q. What is the chance that two people in a group will have the same birth-date: month and day?

A. One chance in 365?

The “Birthday” Problem

Richard von Mises (1883-1953)

In a group of 28 people, a birthday match is “expected”.

The "Birthday" Problem
Math Answer

If the chance of an rare event is *p* and *p* =1/*k*, then this event is "expected" in k trials.

In a group of size N, there are (N-1)(N/2) pairs.

Solve for N(k). k = (N-1)(N/2) = (N^2-N)/2

Quadratic: N^2 – N - 2k = 0

Estimate: ~ N^2/2 = 1/p.

Trial and error: [27^2]/2 = 364 = 1/p = k

Q. Are students convinced? No!!!

49 Connections: Quadrant 1

Schield (2011)			RICHARD VON MISES' BIRTHDAY PROBLEM								28 People	
		Month	10	11	11	9	4	7	6			
		Day	16	18	8	9	13	25	24			
Month	Day										Month	Day
8	20							1			7	25
10	29										8	16
4	11										11	6
3	3										11	29
1	3										8	3
3	30										3	24
10	28										1	15
		Month	5	2	6	2	1	7	5			
		Day	28	8	6	12	14	1	25			

49 Connections: Quadrant 2

Schield (2011)			RICHARD VON MISES' BIRTHDAY PROBLEM								28 People
		Month	8	12	7	11	6	4	2		
		Day	28	2	15	15	5	24	2		
Month	Day									Month	Day
10	8									2	5
5	17									2	17
9	13									12	26
11	18									3	6
12	21							2		4	20
2	28									10	2
10	11									3	23
		Month	10	7	4	12	8	4	8		
		Day	22	22	10	6	4	20	21		

49 Connections: Quadrant 3

Schield (2011)			RICHARD VON MISES' BIRTHDAY PROBLEM							28 People	
		Month	3	8	7	5	6	8	11		
		Day	4	5	25	27	19	4	26		
Month	Day									Month	Day
7	15									12	13
4	31									7	30
11	3									2	1
8	15									4	14
3	28									10	25
3	18									1	18
2	26		3							12	23
		Month	2	3	2	4	6	11	9		
		Day	26	26	23	6	30	11	8		

49 Connections: Quadrant 4

Schield (2011)			RICHARD VON MISES' BIRTHDAY PROBLEM								28 People	
		Month	11	11	3	5	1	5	2			
		Day	5	27	17	3	5	19	4			
Month	Day										Month	Day
11	5		4								11	12
11	17										8	24
8	2										5	1
4	26										3	28
4	22										10	13
10	8										4	4
12	22										8	11
		Month	1	7	5	5	12	10	5			
		Day	2	1	23	7	20	14	14			

Connections: Quadrants 1 & 2

Schield (2011)			RICHARD VON MISES' BIRTHDAY PROBLEM								28 People	
		Month	5	1	3	2	9	10	9			
		Day	24	3	11	22	10	19	31			
Month	Day										Month	Day
5	29										7	2
12	13										2	20
2	17							2			10	9
6	8										5	11
5	28										4	9
7	24						1				9	10
3	28										6	28
		Month	11	6	4	9	3	10	4			
		Day	3	1	12	4	17	9	8			

49 Connections: Top-to-Bottom

Schield (2011) **RICHARD VON MISES' BIRTHDAY PROBLEM** **28 People**

		Month	*11*	*8*	*10*	*10*	*8*	*10*	*3*		
		Day	*19*	*3*	*28*	*17*	*27*	*29*	*5*		
Month	Day					S				Month	Day
5	*23*									*1*	*12*
1	*1*									*11*	*17*
9	*6*									*12*	*3*
10	*13*									*7*	*29*
7	*14*									*2*	*17*
8	*30*									*4*	*2*
1	*8*									*8*	*17*
					N						
		Month	*12*	*3*	*10*	*9*	*12*	*9*	*5*		
		Day	*24*	*6*	*17*	*19*	*1*	*20*	*29*		

49 Connections: Side-To-Side

Schield (2011)			RICHARD VON MISES' BIRTHDAY PROBLEM									28 People
		Month	2	3	10	6	6	9	6			
		Day	14	3	13	27	13	7	24			
Month	Day										Month	Day
1	24										1	31
9	8	E									6	28
12	6										12	24
12	28										10	1
10	27										11	19
9	18									W	9	8
4	12										4	16
		Month	8	8	6	5	7	4	7			
		Day	13	3	19	3	30	9	18			

21 Connections: Same-Side

Schield (2011)			RICHARD VON MISES' BIRTHDAY PROBLEM								28 People	
		Month	3	2	2	3	9	3	5			
		Day	4	5	9	29	20	5	20			
Month	Day										Month	Day
6	22									E	4	1
10	8										7	10
5	5										3	26
11	23										3	10
3	27									E	4	1
10	2										9	8
2	21										5	7
		Month	8	1	10	12	9	5	5			
		Day	18	6	11	9	3	26	19			

Multiple Connections

Schield (2011)			RICHARD VON MISES' BIRTHDAY PROBLEM								28 People	
		Month	11	7	5	7	12	6	9			
		Day	9	11	20	22	15	5	28			
Month	Day			S							Month	Day
1	17										12	17
10	9										7	6
8	29			1					2		7	11
3	24										4	21
2	9										5	24
6	29										1	14
11	12										3	12
									N			
		Month	11	12	9	12	4	4	7			
		Day	27	29	21	14	16	10	11			

Multiple Connections

Schield (2011)			RICHARD VON MISES' BIRTHDAY PROBLEM								28 People	
		Month	2	10	5	10	10	7	7			
		Day	19	12	12	20	17	20	5			
Month	Day			N		N					Month	Day
10	4			2							9	21
4	3					1					10	20
9	31										6	15
6	1									E	11	7
5	23									E	11	7
5	23										4	2
9	2										6	12
		Month	4	9	7	1	9	7	4			
		Day	12	21	12	19	16	4	27			

Connections and Chance

Pairs	GROUP	Details
196	Quadrants 1-4	49 pairs each
49	Side-to-Side	
49	Top-to-Bottom	
84	Within each side	21 pairs each
378	TOTAL	

A “birthday” match has one chance in 365.

In a group of 28, we have 378 pairs: (N-1)(N/2).

A match is expected: Match is more likely than not.

Conclusion

Students need to “see” that coincidences

1. are more common than expected

2. depend on the context

3. compare ex-ante with ex-post

4. may still be signs of causation (Cholera)

That runs with 1 chance in N are generally found in N tries.

ASA Chapter & StatChat
May 9 Wednesday 6 PM Augsburg

Wed May 9. Augsburg College. 6-9 PM. Supper

Chapter website: **www.amstat.org/chapters/twincities/**

SPEAKERS:

Marc Isaacson: **Teaching Activities**

Robert Raymond: **Untangling a Conundrum.**

Milo Schield: **Introducing the Matrixx Case**

Danny Kaplan: **Comments on US Supreme Court Matrixx Case: Is Significance Significant?**