

May 2012 1

Big Data Creates Beguiling Coincidences

MILO SCHIELD,
Augsburg College
Director, W. M. Keck Statistical Literacy Project
US Rep, International Statistical Literacy Project
Member, International Statistical Institute

May, 2012
These slides:
www.statlit.org/pdf/2012Schield3Keene6up.pdf
Related paper at www.StatLit.org/pdf/2012Schield-MAA.pdf

May 2012 2

Coincidence?



Coincident or Not?

If,
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Equals,
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26


Then,
 $K + N + O + W + L + E + D + G + E = 96\%$
 $H + A + R + D + W + O + R + K = 98\%$

Both are important, but the total falls just short of 100%

But,
 $A + T + T + I + T + U + D + E = 100\%$

May 2012 4

Coincidence?



- Love
- Money
- Astrology
- Tarot

[submit](#)

SpiritNow.com

May 2012 5

Not Coincidence

The two major products that came out of Berkeley:
LSD and UNIX.

We don't believe this is coincidence!

Computer programmer: Jeremy S. Anderson

May 2012 6

Coincidence?

Virginia Woman Wins \$1 Million Lottery Twice in the Same Day

By CHRISTINA NG | Good Morning America – Mon, Apr 23, 2012

[Email](#) | [Recommend](#) | 214 | [Tweet](#) | 19 | [Share](#) | 8 | [Print](#)

RELATED CONTENT

Winning the lottery once in a lifetime is pretty lucky. [Winning the lottery twice in the same day?](#) Virginia Pike is one of the few people that can describe that feeling.

The Berryville, Va., resident had two tickets that matched five of the six Powerball numbers in an April 7 drawing so that each ticket was worth \$1 million.

Virginia Woman Wins \$1 Million ...

"I'm in shock!" Fike said in a news release from the Virginia lottery.

Cancer Cluster In St. Paul: Correlation Or Coincidence?

May 2, 2012 10:54 PM

MINNEAPOLIS (WCCO) – Residents of a [St. Paul](#) neighborhood with higher than average cancer cases are wondering if it could be linked to nearby hazardous waste sites.

“higher than average...”



May 2012

Run of Heads (Red Cells): Chance of 5 Touching: 1 in 32 (2^5)

B4 =RANDBETWEEN(0,1)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	A	A		
1	Fair coin: find longest run of heads in a row																														
2	5	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0
3										1						2															
4	3	1	1	0	1	1	0	1	1	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
5																															
6	4	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	1
7																															
8	5	0	1	1	1	1	0	0	0	1	1	0	1	1	0	0	0	1	1	0	1	1	0	1	1	0	1	1	0	1	

www.StatLit.org/Excel/2012Schield-Runs.xls

May 2012

Run of 10 heads (Ones): Chance = 1 in 1,024 (2^10)

B4 =RANDBETWEEN(0,1)

Fair coin: Longest run of heads (red cells) in a row

www.StatLit.org/Excel/2012Schield-Runs.xls

May 2012

Clusters: Grains of Rice Chance of Red: One in 10

A3 =RANDBETWEEN(0,9)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
3	9	3	9	9	9	9	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

www.StatLit.org/Excel/2012Schield-Rice.xls

May 2012

Clusters: Grains of Rice Five cells touch: 1 chance in 100,000

A3 =RANDBETWEEN(0,9)

Rice-10 sheet

Find the largest group of high cells (red fill) that are touching each other.

a. touching on sides in a row b. touching on sides (but not just on points)

www.StatLit.org/Excel/2012Schield-Rice.xls

May 2012 12

Runs and clusters are much more likely than expected!

When students press F9, they often get:


RUNS	CLUSTERS
a run of 10 heads: one chance in 2^10	a cluster of six squares: one chance in 10^6
a “thousand-year flood” every year	a “million-year flood” every year

They get unlikely results every time!

May 2012 13

Coincidence is much more likely than expected!!!

WHY?
HOW?



May 2012 14

**Explanation #1
The question is ambiguous**

What is the chance of “that”?

- At a specific place or anywhere?
- Paint the target before or after the shooting?
- Before or after the fact (ex post vs. ex ante)?

Consider a run of 8 heads:

- One chance in 256 **at a pre-designated spot**
- Close to 50% **somewhere** in the next 256 flips

May 2012 15


Explanation #2

Law of Very-Large Numbers:
‘Impossible’: almost certain given enough tries

The Law of Very-Large Numbers isn’t covered in most intro-stats courses where the focus is on sampling error in small samples.


May 2012 16

#2 Law of Very Large Numbers with Compression/Overlap



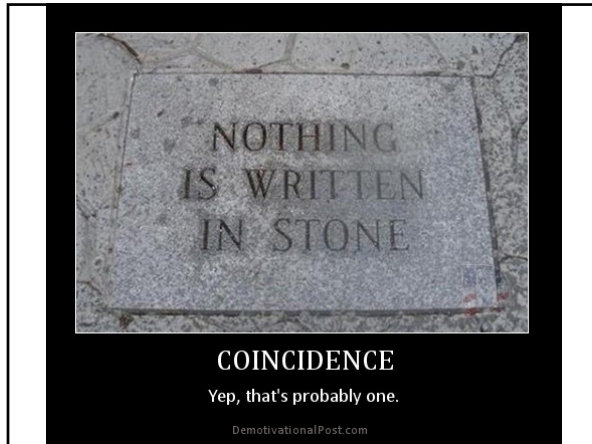
May 2012 17

In ancient times, just a coincidence?

COINCIDENCE

Because you couldn't have planned it any better.



May 2012 21

Coincidence vs. Meaning

the POWER of COINCIDENCE
THE MYSTERIOUS ROLE OF SYNCHRONICITY IN SHAPING OUR LIVES
FRANK JOSEPH

A Devil's Book of COINCIDENCE
John Hartman

The THIRD Coincidence
DAVID BISHOP

The Power of COINCIDENCE
How Life Shows Us What We Need to Know

May 2012 22

Coincidence vs. Meaning

BEYOND COINCIDENCE
AMAZING STORIES OF COINCIDENCE AND THE MYSTERY BEHIND THEM
MARTIN PLIMMER AND BRIAN KING

@JTalarico328
John Talarico
May 1st 1945 Hitler confirmed dead. May 1st 2011, Bin Laden confirmed dead.
3 hours ago via web · Favorite · Retweet · Reply
Retweeted by MrBerro and others

May 2012 23

Coincidence vs. Meaning

Chuck Norris was born on May 6th, 1945

Nazi Germany surrendered on May 7th, 1945

COINCIDENCE?
I think not!

Born 6.May.1945

Capitulated 7.May.1945

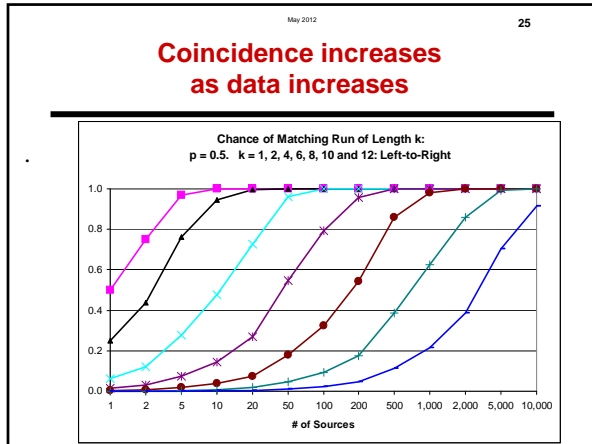
Coincidence?
Don't think so

May 2012 24

Coincidence increases as data increases

P = chance of success on next try
 K = length of a run of successes
 Chance of a run of length k : p^k . Decreases as $K \uparrow$

n = number of independent factors
 Chance of two matching runs of length K :
 $1 - (1 - P^K)^n$. Increases as $n \uparrow$



May 2012 26

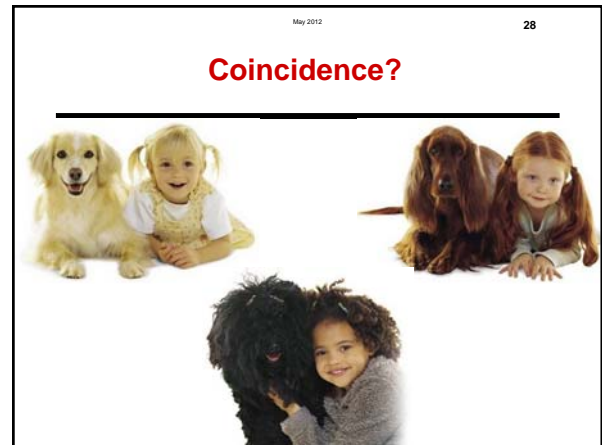
Jung's theory of Synchronicity

Synchronicity: “ ‘coincidences’ which were connected so meaningfully that their ‘chance’ concurrence would represent a degree of improbability that would have to be expressed by an astronomical figure.”

Principle: “Everything in the universe has some kind of correspondence one to the other.”

May 2012 27

Coincidence Today



- May 2012 30
- ### Students must see that “coincidence”:
1. is much more common than expected
 2. may involve an ambiguity (ex-ante/ex-post)
 3. may involve Law of Very-Large Numbers.
Unlikely is almost certain given enough tries
 4. may still be a sign of causation.
1854 cholera outbreak in London at a pump.