

Change Index

A New Statistic?

Kirke Bent

kirke.bent@verizon.net

973-216-9652

Change Index

This talk is about a new or at least unfamiliar statistic called Change Index.

Change Index is a measurement of how much a “report “has changed since the previous edition.

It measures heterogeneous as well as homogeneous reports.

Why detecting significant change is hard

Change Index

Significance

Calculating CI

What's the problem?

Let's say every week we get a report, perhaps a financial report or a call center operational report. We get LOTS of other reports, too.

Q: How do we decide whether to dig in to *this* report?

A: When there is a significant change from a prior version.

Why is it hard?

- + There may be too many lines to take in.
- + Many small changes may obscure the picture.
- + Significance is subjective. Judgments *will* differ.
- + Conventional statistics don't work well here.
- + We must relate different categories to get an overall view.

Why detecting change is hard

Change Index

Significance

Calculating CI

Change Index - Real data

US Dry Edible Beans

US Dry Edible Beans	2011	2012	[8 more]
Black Acres Harvested	177,900	212,800	
Black Acres Planted	185,900	216,800	
Black Production CWT	3,018,000	3,739,000	
Black Yield LB/Acre	1,696	1,757	
Baby Lima Acres Harvested	12,500	12,600	
Baby Lima Acres Planted	12,600	12,900	
[18 More]			

Source: www.usda.gov

All the Data

US Dry Edible Bean Production	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003
BLACK ACRES HARVESTED	212800	177900	278300	179300	168900	171600	159300	107100	127600	80700
BLACK ACRES PLANTED	216800	183900	284000	187400	171900	175700	167400	111600	138300	84300
BLACK PRODUCTION CWT	3739000	3018000	4661000	3010000	2923000	2803000	2673000	1811000	1892000	1263000
BLACK YIELD LB / ACRE	1757	1696	1675	1679	1731	1633	1678	1691	1483	1565
BLACK EYE ACRES HARVESTED	32800	23300	32700	45400	27300	27100	29400	21500	25300	46100
BLACK EYE ACRES PLANTED	37400	24600	34700	48300	29300	27800	31400	23000	28000	50500
BLACK EYE PRODUCTION CWT	546000	337000	585000	771000	394000	497000	533000	406000	384000	785000
BLACK EYE YIELD LB / ACRE	1665	1446	1789	1698	1443	1834	1813	1888	1518	1703
DARK RED KIDNEY ACRES HARVESTED	44700	47700	45700	47300	49300	39100	46400	58000	46600	48400
DARK RED KIDNEY ACRES PLANTED	46200	48800	48500	50500	50800	40200	48800	60700	51300	49900
DARK RED KIDNEY PRODUCTION CWT	905000	790000	833000	850000	992000	663000	824000	1048000	686000	845000
DARK RED KIDNEY YIELD LB / ACRE	2025	1656	1823	1797	2012	1696	1776	1807	1472	1746
GREAT NORTHERN ACRES HARVESTED	52100	59700	69900	48300	71100	57000	59300	71200	46800	103800
GREAT NORTHERN ACRES PLANTED	55600	61800	78500	53900	76100	59500	69700	72800	51100	109400
GREAT NORTHERN PRODUCTION CWT	1222000	1196000	1403000	999000	1598000	1186000	1190000	1585000	951000	2216000
GREAT NORTHERN YIELD LB / ACRE	2345	2003	2007	2068	2248	2081	2007	2226	2032	2135
LIGHT RED KIDNEY ACRES HARVESTED	38900	35700	49400	52400	54200	46000	40800	68800	51500	64100
LIGHT RED KIDNEY ACRES PLANTED	40200	37200	53100	56300	56300	47400	44400	71400	54700	67100
LIGHT RED KIDNEY PRODUCTION CWT	802000	642000	966000	967000	1023000	813000	770000	1109000	816000	1095000
LIGHT RED KIDNEY YIELD LB / ACRE	2062	1798	1955	1845	1887	1767	1887	1612	1584	1708
BABY LIMA ACRES HARVESTED	12600	12500	12200	14600	11700	15600	13000	16400	10900	14100
BABY LIMA ACRES PLANTED	12900	12600	12200	15200	11700	16000	13500	16700	11300	14500
BABY LIMA PRODUCTION CWT	306000	236000	304000	352000	239000	377000	304000	385000	267000	325000
BABY LIMA YIELD LB / ACRE	2430	1890	2490	2410	2040	2420	2340	2350	2450	2300
Source: WWW.USDA.GOV										


```
B11v12R.csv - Notepad
File Edit Format View Help
BLACK - ACRES HARVESTED,"177,900","212,800"
BLACK - ACRES PLANTED,"183,900","216,800"
BLACK - PRODUCTION IN CWT,"3,018,000","3,739,000"
BLACK - YIELD IN LB / ACRE,"1,696","1,757"
BLACK EYE - ACRES HARVESTED,"23,300","32,800"
BLACK EYE - ACRES PLANTED,"24,600","37,400"
BLACK EYE - PRODUCTION IN CWT,"337,000","546,000"
BLACK EYE - YIELD IN LB / ACRE,"1,446","1,665"
DARK RED KIDNEY - ACRES HARVESTED,"47,700","44,700"
DARK RED KIDNEY - ACRES PLANTED,"48,800","46,200"
DARK RED KIDNEY - PRODUCTION IN CWT,"790,000","905,000"
DARK RED KIDNEY - YIELD IN LB / ACRE,"1,656","2,025"
GREAT NORTHERN - ACRES HARVESTED,"59,700","52,100"
GREAT NORTHERN - ACRES PLANTED,"61,800","55,600"
GREAT NORTHERN - PRODUCTION IN CWT,"1,196,000","1,222,000"
GREAT NORTHERN - YIELD IN LB / ACRE,"2,003","2,345"
LIGHT RED KIDNEY - ACRES HARVESTED,"35,700","38,900"
LIGHT RED KIDNEY - ACRES PLANTED,"37,200","40,200"
"LIGHT RED KIDNEY - PRODUCTION, IN CWT","642,000","802,000"
"LIGHT RED KIDNEY - YIELD, IN LB / ACRE","1,798","2,062"
BABY LIMA - ACRES HARVESTED,"12,500","12,600"
```

Actual input format

Just the top line of the previous slide

The format is CSV:

name, base [2011], new [2012]

BLACK - ACRES HARVESTED, "177,900", "212,800"

Change Index (CI) Calculator

Set file locations.

Paths may be absolute or relative to the current location.

Enter or accept input file name (with path if needed).

No header line, please.

Enter or accept output file name (with path if needed).

Enter or accept ID of this run

Click to accept current values.

CI: First Screen - Identify Files

Change Index (CI) Calculator

Calculated Results

0.23 = CI (Change Index)

0.24 = RCI (Relative Change Index)

Top Contributors to CI

Percent	Cum %	Name
85.135%	85.135%	BLACK - PRODUCTION IN CWT
7.154%	92.288%	BLACKEYE - PRODUCTION IN CWT
4.193%	96.481%	LIGHT RED KIDNEY - PRODUCTION IN CWT
2.166%	98.647%	DARK RED KIDNEY - PRODUCTION IN CWT
0.802%	99.449%	BABY LIMA - PRODUCTION IN CWT

Counts

24 Number of categories

0 Number added

0 Number dropped

CI: Second Screen - Calculated results

	A	B	C	D	E	F	G	H	
1	CI Calculation - Run ID= Mon May 20 2013 09:54								
2	Calculated Result								
3	0.227229	CI (Change Index)							
4	0.205245	DI (Difference Index)							
5	0.23939	RCI (Relative Change Index)							
6									
7	Contribution to CI, RCI								
8	% CI Contr	Cum %	Name	% RCI Contrib					
9	85.13477	85.13477	BLACK - A	0.199474	0.227229				
10	7.153672	92.28845	BLACK - A	0.177267	0.227229				
11	---	---	(Many lines omitted.)						
12	7.85E-06	100	BABY LIMA	1.47E-05	0.227229				
13	1.64E-06	100	BABY LIMA	0.802477	0.227229				
14	6.09E-07	100	BABY LIMA	4.78E-05	0.227229				
15									
16	Counts								
17	24	Number of categories							
18	0	Number added							
19	0	Number dropped							
20									
21	Summary Statistics								
22	0.009468	Mean	CI Contribution						
23	3.58E-06	Median	CI Contribution						

Detail output file, opened in Excel

Why detecting change is hard

Change Index

Significance

Calculating CI

What's a "significant" change?

- + It all depends.
- + It depends on context, frequently historical.
- + It's a rare number that has meaning without context.

38¢ \Rightarrow 21¢

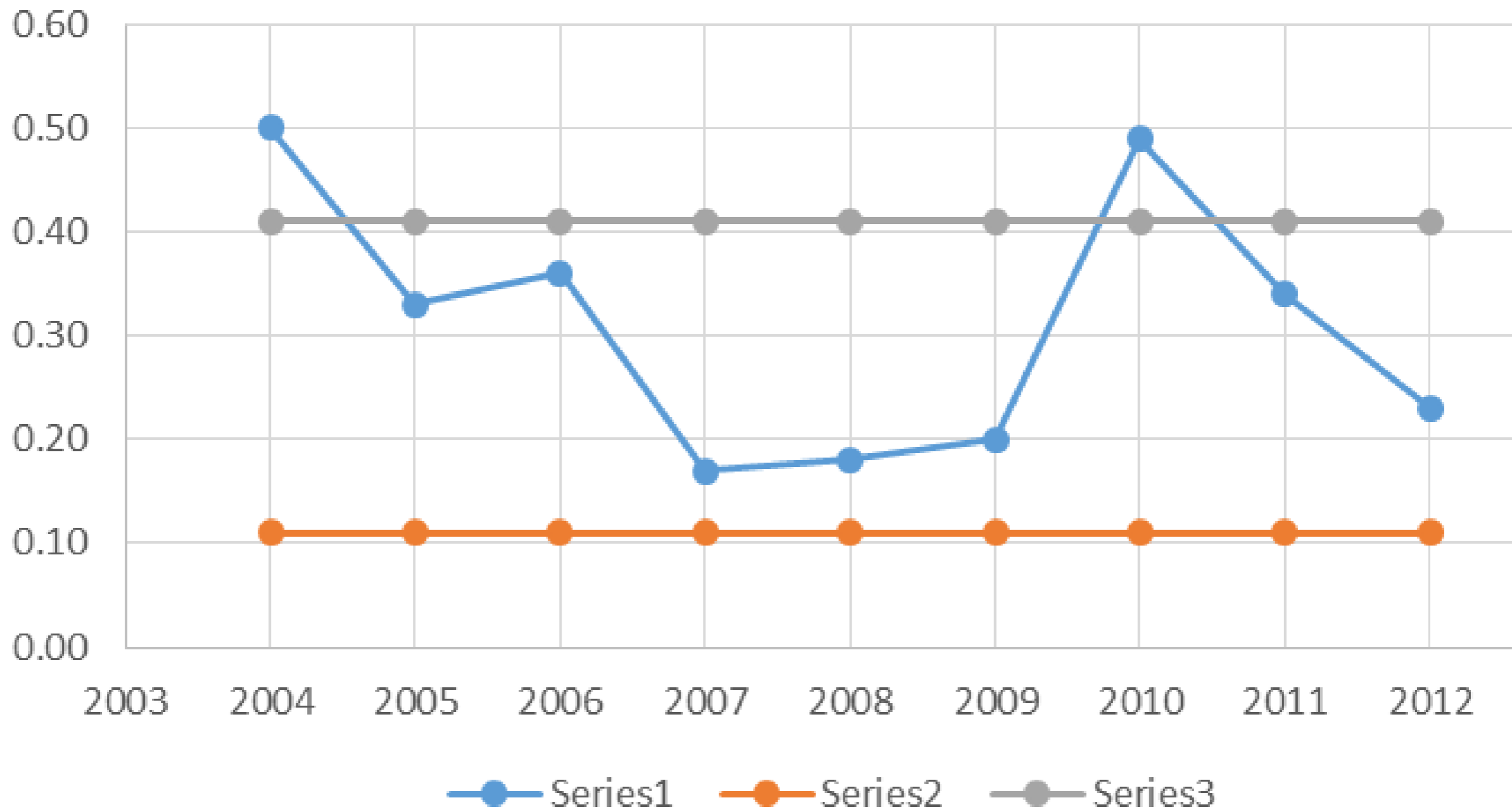
2.7% \Rightarrow 3.2%

40.0%

How do we get a handle on
CI significance?

Comparison

Actual CI 2004-2012



Why detecting change is hard

Change Index

Significance

Calculating CI

How CI is defined

- + We work with three columns of numbers (or “vectors”).
- + They are called: the Base figures, the New, and the calculated Difference .
- + We calculate a “measure” for each column.
- + The CI is the ratio of the Difference measure to the Base measure.

How the “measure” is calculated

The calculation is easy:

- +Square each element in a column.

- +Sum the squares.

- +Take the square root of the sum.

Consequences of this definition

- + All lines participate in the calculation.
- + Negative changes do not offset.
- + The effect of smaller changes is minimized.
- + CI is objective, quantified, and repeatable

Why not Management by Exception?

We can roughly define this as picking a key measurement, say Sales, and only paying attention if this falls by a chosen amount.

Feedback, Please

- + Is this really a problem?
- + Does CI look like a good approach?
- + Problems with CI?
- + What are the alternatives?

If anyone would like to Beta Test this on their data, I'd be glad to help them get started.

If you're interested, please give me your card or contact me.

kirke.bent@verizon.net 973-216-9652

Thanks

Predictive

"Don't rely on past data when you can predict what will happen tomorrow!" web ad for 9/25/13 SAP seminar.

Causation

21

$38\phi \Rightarrow 21\phi$

Analysts were cheered.

$38\text{¢} \Rightarrow 21\text{¢}$

$2.7\% \Rightarrow 3.2\%$

Why it works that way

- + Negative changes do not offset.

 - The square of a negative is positive.

- + The effect of smaller changes is minimized.

 - Square is not linear.

Change Index is a measure of change.