

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2

Mathematics for Computer Science
MIT 6.042J/18.062J

Cherry Picking voids Confidence



Albert R Meyer, April 29, 2016

select.1

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2

Mathematics for Computer Science
MIT 6.042J/18.062J

Selective Sampling voids Confidence



Albert R Meyer, April 29, 2016

select.2

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2

TB LAB test

Lab offers alternative TB test that managers believe is 95% accurate.

It's independent of prior test, so useful to confirm diagnoses.



Albert R Meyer, April 29, 2016

select.3

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2

TB LAB test

Since prior test is 99% correct, managers think their retest should show about 1% errors. (Actually $1 \pm (0.05)\%$).



Albert R Meyer, April 29, 2016

select.4

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2

TB LAB test

Since prior test is 99% correct, managers think their retest should show about 1% errors. They are upset when **nearly all their retests disagree with the prior test.**



Albert R Meyer, April 29, 2016

select.5

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2

TB LAB test broken?

Should they be upset? **No!**
Is their test broken? **No!**
So what's wrong with their reasoning?



Albert R Meyer, April 29, 2016

select.6

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2

99% Confident Sample

If the lab got **random sample** of prior test results **from everyone**



Albert R Meyer, April 29, 2016

select.7

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2

99% Confident Sample

If the lab got **random sample** of prior test results **from everyone**, then they should indeed expect 99% prior tests correct. But they get **selective sample**: people who opted for their test.



Albert R Meyer, April 29, 2016

select.8

6	9	13	7
12	10	5	
3	1	4	14
15	8	11	2

Cherry Picking voids Confidence

Confidence is
 $\Pr[\text{mistake over all trials}]$
 Not same as
 $\Pr[\text{mistake over select trials}]$

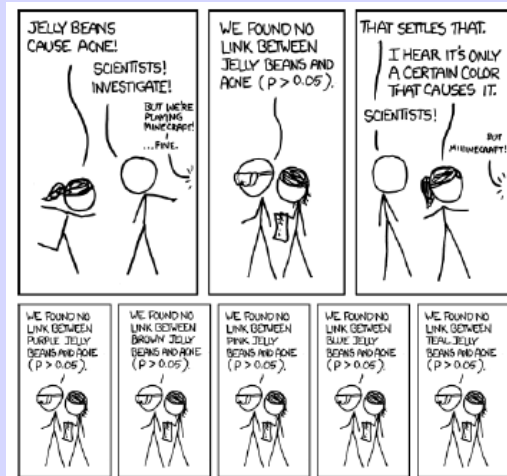


Albert R Meyer, April 29, 2016

select.9

<http://xkcd.com/882/>

6	9	13	7
12	10	5	
3	1	4	14
15	8	11	2

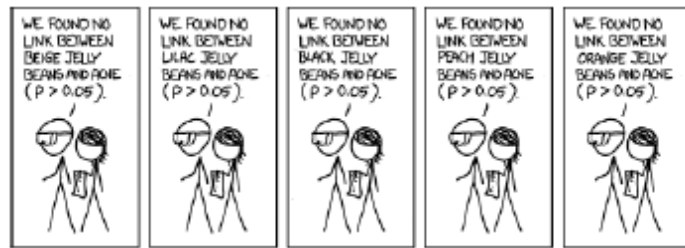


Albert R Meyer, April 29, 2016

select.10

<http://xkcd.com/882/>

6	9	13	7
12	10	5	
3	1	4	14
15	8	11	2

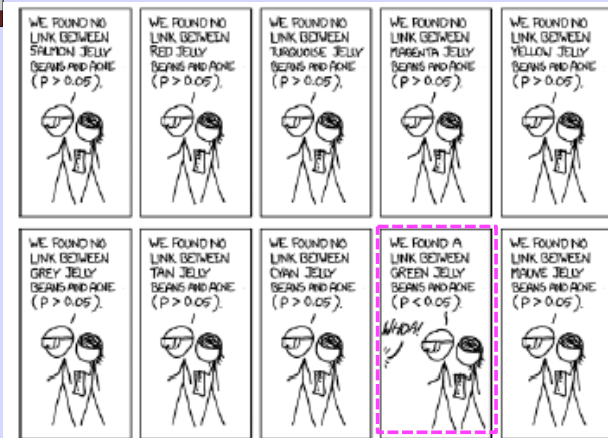


Albert R Meyer, April 29, 2016

select.11

<http://xkcd.com/882/>

6	9	13	7
12	10	5	
3	1	4	14
15	8	11	2

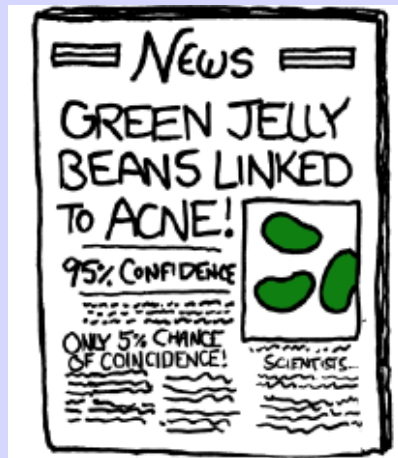


Albert R Meyer, April 29, 2016

select.12

<http://xkcd.com/882/>

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2



Albert R Meyer, April 29, 2016

select.13

Cherry Picking voids Confidence

Do 20 tests with 95% confidence, but only report the one that shows positive. Silly to assert confidence in that one report.

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2



Albert R Meyer, April 29, 2016

select.14

Selective Sample

If the lab got **random sample** of prior test results **from everyone**, then they should indeed expect 99% prior correct.

But their **retest sample** is from worried people with **TB+ prior test result**.

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2



Albert R Meyer, April 29, 2016

select.15

Selective Sample

Lab managers expected **confidence** error rate,
 $\Pr[\text{prior test mistake}]$

but got
 $\Pr[\text{prior mistake} | \text{retested}]$

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2



Albert R Meyer, April 29, 2016

select.18

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2

Selective Sample

Lab managers expected
confidence error rate,

$$\Pr[+ | \text{no TB}] = 0.01$$

but got

$$\Pr[\text{prior mistake} | \text{retested}]$$



Albert R Meyer, April 29, 2016

select.19

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2

Selective Sample

Lab managers expected
confidence error rate,

$$\Pr[+ | \text{no TB}] = 0.01$$

but got **predictive probability**

$$\Pr[\text{no TB} | +] \approx 0.99$$



Albert R Meyer, April 29, 2016

select.20

6	9	13	7
12		10	5
3	1	4	14
15	8	11	2

Cherry Picking voids Confidence

Moral: For claimed confidence
ask "Why am I hearing about
this particular outcome?
How many others were tried
and not reported?"



Albert R Meyer, April 29, 2016

select.21