Which data should be Benford?

Which data should not be Benford?

ANSWER:

Big Order of Magnitude usually implies Benford,

and also…

The histogram should be falling to the right, so that "small is beautiful".

Uniform is flat hence it's anti-Benford!
Normal is symmetrical hence it's anti-Benford!

But the fall, in the aggregate, overall, should be calibrated the Benford-way,

not falling too sharply
not falling too gently

Order Of Magnitude = LOG[Max/Min]

It should be at least **3**

In other words:
Data should have a lot of variability, large spread.

But is order of magnitude a **robust** measure of variability?

NO!

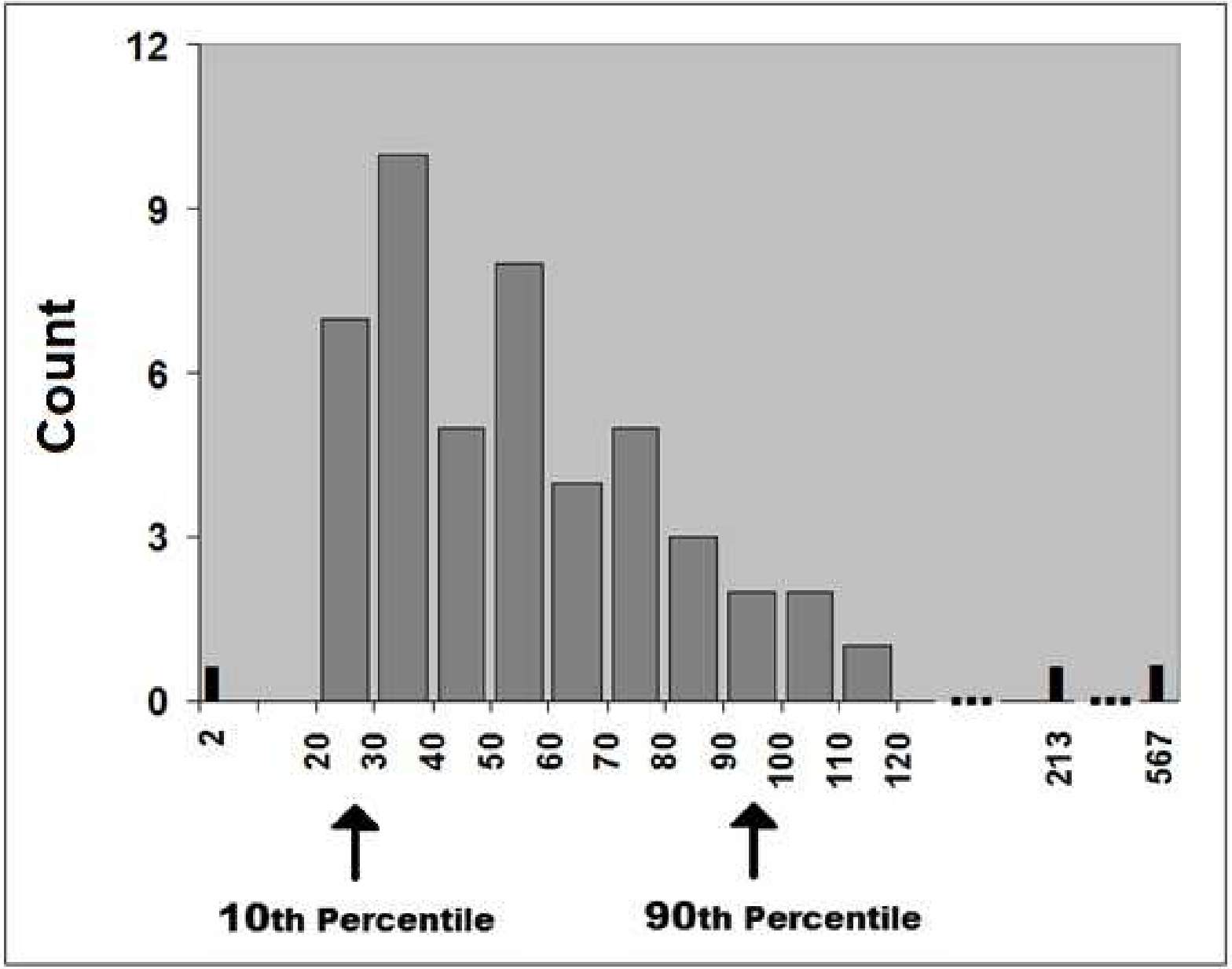
Deceptive wide range on the x-axis occurs when outliers are included.

It is necessary to **ELIMINATE** the outliers and whiskers before we take measurement!

Cut those whiskers off!

For example, data is:

2	23	24	25	26	27	28	29	32	33
33	33	34	36	37	38	38	39	40	41
42	47	48	50	51	52	53	55	56	57
59	60	63	67	68	75	76	77	78	79
80	84	86	91	94	103	107	114	**213**	**567**


Count
12
9
6
3
0
2
20
30
40
50
60
70
80
90
100
110
120
213
567
10th Percentile
90th Percentile

Naïve Order of Magnitude

Order Of Magnitude = LOG[Max/Min]

Order Of Magnitude = LOG[567/2]

Order Of Magnitude = LOG[284]

Naive Order Of Magnitude = 2.45

Oh… so almost Benford ! Great!

Realistic & Authentic Order of Magnitude

Order Of Magnitude = LOG[~~Max/Min~~]

Order Of Magnitude = LOG[$Q_{90\%}/Q_{10\%}$]

Order Of Magnitude = LOG[94.9/26.9]

Order Of Magnitude = LOG[3.53]

Realistic Order Of Magnitude = 0.55

So no, it's not Benford! Sorry!

A robust measure of O.O.M.

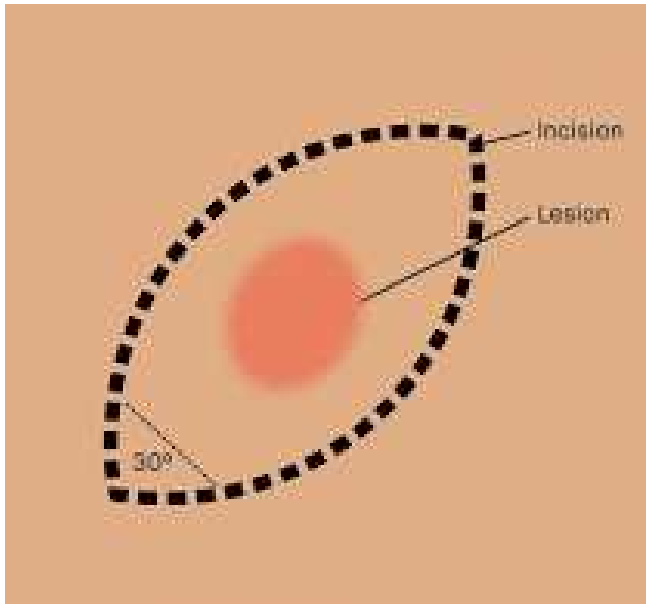
Core Order of Magnitude =

$$Q_{90\%}/Q_{10\%}$$

Core Order of Magnitude =

90th percentile / 10th percentile

Skin Cancer



Is the surgeon doctor cutting too much around the tumor even in the healthy tissue – in order to ensure that the cancer does not return?

A more conservative trim:

Core Order of Magnitude =

$Q_{95\%}/Q_{5\%}$

Core Order of Magnitude =

95th percentile / 5th percentile

Perhaps Professor **Marco Riani** could suggest a trimming algorithm, constructing some more innovative robustness methods for Benford.

We need a solid rule of thumb here!

END