

In God we trust. Everyone else must bring data.

The word data is the plural form of its Latin root *datum*, which means ‘something given’. Agencies are paid money for adding value, so we can’t just pass on the ‘given’ data to the consumer – we have to add something.

The word hyperbole might not have come to mind, but that’s what we add to break through the clutter – good old hype (which I once saw so aptly defined as: ‘an extension of the truth’).

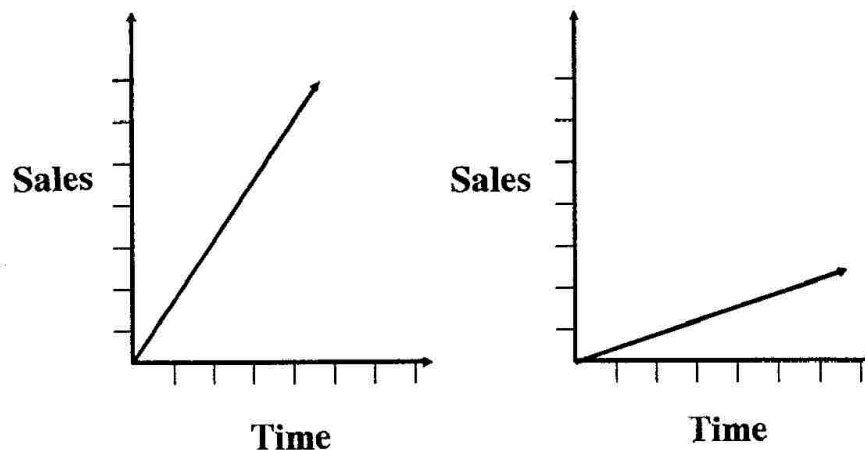
Statistics seem to confer a saintliness on data. We tend to hold the belief that numbers don’t lie, so when we see the following, we are shocked at this ‘fact’:

“The average age of youngsters getting into drugs today is 11 years old”.

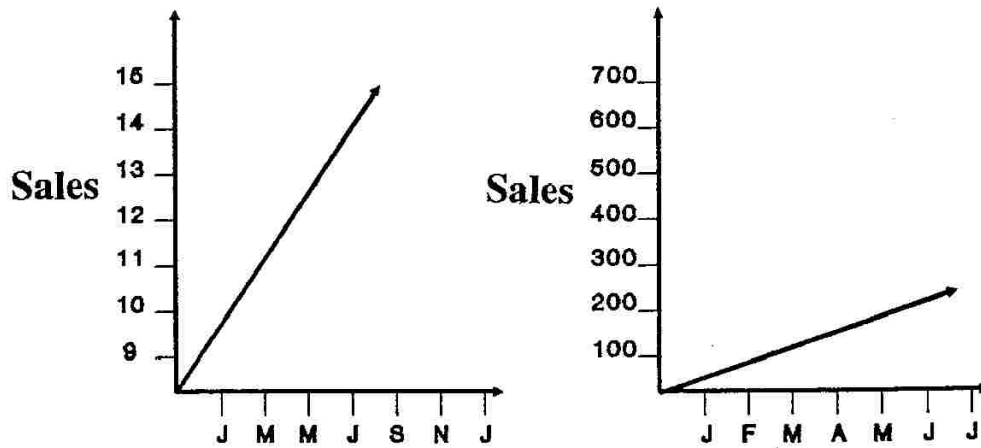
But this entire sentence has absolutely no statistical merit. ‘Average’ of what? Youngsters you might say. But what is a youngster? It all depends on the sample. I could prove all youngsters getting into drugs are 9 years old. Or 8 or 6 – it just depends on which ages I classify as youngsters. And what are drugs? I used to love building model airplanes when I was a kid. My parents also loved it – it was a healthy pursuit that kept me off the streets. The fact that I used an inordinate amount of glue did not seem to phase them. Drugs are in the nose of the beholder.

But numbers come into their own when we need to reflect more as most and a little as a lot.

Take a quick look at the following two graphs. Which one shows better growth?

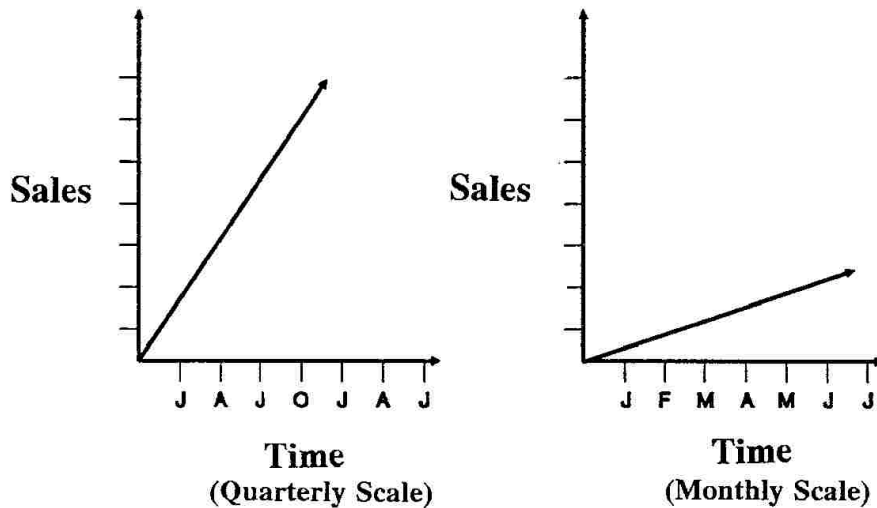


Now let's include the scale, and the whole story changes:

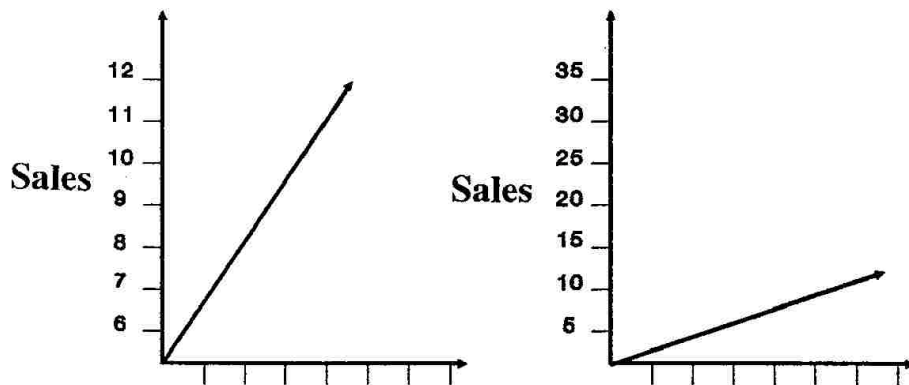


This technique is great for investment companies – all they need to do is get the line as steep as possible by adjusting time and space:

Firstly, they can compress the time scale on the horizontal axis



Or by exaggerating the scale on the Y-axis by only using a part of it (and not starting at zero).



There is, of course, a third way to increase the slope of the graph – just perform better, but that's like really hard work.

If a picture is really worth a thousand words, then we can make two pictures worth four thousand words. Mark my words:

Let me represent my savings of R100 with a line 1 cm in length. The bank tells me that they can double my money in five years, which I can reflect pictorially on a line double that length:

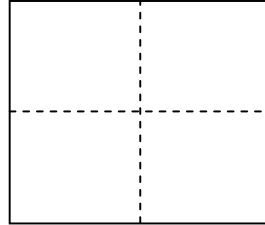
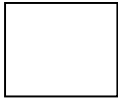
My R100 before



My R100 after five years (now R200)



But let's get creative and represent the line as a square (1 cm by 1 cm). They can then show my savings after the 'doubling' as a square 2 cm by 2 cm.



But, as you can see with the dotted lines, a doubling of dimensions results in four times the space.

Let's say they get really creative and want to show the 'doubling' by means of a piggybank:



Anyone can check – the piggy on the right is twice as high and twice as broad as the one on the left, so it's twice as big. But to us enlightened few, it's actually four times as big.

The diet industry also serves up some interesting stats. Although "I lost 15 kg in 20 days" cannot be disputed, how does one lose 20 inches? Sure, I can lose 4 inches off my waist, but I can actually lose about a kilometer if I measured myself in enough places.

If the above seems a bit simplistic, chew on this; Imagine you are being tested for a disease that is prevalent in 0.5% of the population. The test is 98% accurate. If it came back positive, how sure would you be of having that disease?

- a) 98%
- b) 50%
- c) About 20%
- d) Can't say

See answer below.

We are all affected by statistics. The day we are born we become one (actually we influence the statistic for pregnant mothers well before that). We then spend a large part of our lives trying to fathom out life, using 'proven facts' to help us along the way. When I told my mom she had a 0.0000001 chance of winning the lotto (and sarcastically added that the lotto is for people who don't understand statistics), she correctly pointed out that if she did not buy a ticket, she would have no chance of winning. I stand corrected.

About the author

Sid Peimer is the middle child of four children. He is the mayor of www.stratplanning.com, the most popular site on the internet (data on file).

Answer: About 20%

Just because the test is 98% accurate, it does not mean that there is a 98% chance you have the disease.

Say we test 10 000 people, then 50 of those people actually have the disease (the prevalence is 0.5%).

98% of these people will test positive, which means that there will be 49 positive tests. Of the 9 550 disease-free people, 2% (or 191) are false positives. There is therefor a total of 240 positive tests (49 correct and 191 incorrect). So the probability of a correct positive test is 49/240 or just over 20% in this case. Not so smart now are you?