Regression Toward the Mean: An Introduction with Examples

https://fs.blog/2015/07/regression-to-the-mean

Regression to the mean is a common statistical phenomenon that can mislead us when we observe the world. Learning to recognize when regression to the mean is at play can help us avoid misinterpreting data and seeing patterns that don't exist.

It is important to minimize instances of bad judgment and address the weak spots in our reasoning. Learning about regression to the mean can help us.

Nobel prize-winning psychologist **Daniel Kahneman** wrote a book about biases that cloud our reasoning and distort our perception of reality. It turns out there is a whole set of logical errors that we commit because our intuition and brains do not deal well with simple statistics. One of the errors that he examines in Thinking Fast and Slow is the infamous regression toward the mean.

The rule goes that, in any series with complex phenomena that are dependent on many variables, where chance is involved, extreme outcomes tend to be followed by more moderate ones.

In Seeking Wisdom, Peter Bevelin offers the example of John, who was dissatisfied with the performance of some of his new employees. Therefore, he put the poor performing employees into a skill-enhancing program where he measured the employees' skill at the beginning of the training program and at the end:

> *Their scores where higher at the end than they were at the beginning. John's conclusion: "The skill-enhancing program caused improvement in skill." This isn't necessarily true. Their higher scores could be the result of*

> *regression to the mean. Since these individuals were initially identified because they were on the low end of the scale of skill, they might have shown an improvement even if they hadn't taken the skill-enhancing program. And there could be many reasons for their earlier performance — stress, fatigue, sickness, distraction, etc. Their true ability perhaps hasn't changed.*
>
> *Our performance always varies around some average true performance. Extreme performance tends to get less extreme the next time. Why? Testing measurements can never be exact. All measurements are made up of one true part and one random error part. When the measurements are extreme, they are likely to be partly caused by chance. Chance is likely to contribute less on the second time we measure performance.*
>
> *If we switch from one way of doing something to another merely because we are unsuccessful, it's very likely that we do better the next time even if the new way of doing something is equal or worse.*

This is one of the reasons it's dangerous to extrapolate from small sample sizes, as the data might not be representative of the distribution. It's also why James March argues that the longer someone stays in their job, "the less the probable difference between the observed record of performance and actual ability." Anything can happen in the short run, especially in any effort that involves a combination of skill and luck. (The ratio of skill to luck also impacts regression to the mean.)

Regression to the Mean

The effects of regression to the mean can frequently be observed in sports, where the effect causes plenty of unjustified speculations.

In Thinking Fast and Slow, Kahneman recalls watching men's ski jump, a discipline where the final score is a combination of two separate jumps. Aware of the regression to the mean, Kahneman was startled to hear the commentator's predictions about the second jump. He writes:

> *"The Norwegian athlete had such a great first jump that for his second jump he will probably be tense, hoping to protect his lead so he will probably do worse" or "The Swedish athlete had a bad first jump and now he knows he has nothing to lose, so he will probably be relaxed, which should help him do better.*

Kahneman points out that the commentator had noticed the regression to the mean and come up with a story for which there was no causal evidence (see narrative fallacy). This is not to say that his story could not be true. Maybe, if we measured the heart rates before each jump, we would see that athletes are more relaxed if their first jump was bad. However, that's not the point. The point is, regression to the mean happens when luck plays a role, as it did in the outcome of the first jump.

The lesson from sports applies to any activity where chance plays a role. We often attach explanations of our influence over a particular process to the progress or lack of it.

In reality, the science of performance is complex, situation dependent and often much of what we think is within our control is truly random.

In the case of ski jumps, a strong wind against the jumper will lead to even the best athlete showing mediocre results. Similarly, a strong wind and ski conditions in favor of a mediocre jumper may lead to a considerable, but a temporary bump in his results. These effects, however, will disappear once the conditions change and the results will regress back to normal.

Regression to the man can have serious implications for coaching and performance tracking. Statistical thinking tells us, that when evaluating performance or hiring, we must rely on track records more than outcomes of specific situations. Otherwise, we are prone to be disappointed.

When Kahneman was giving a lecture to Israeli Air Force about the psychology of effective training, one of the officers shared his experience that extending praise to his subordinates led to worse performance, whereas scolding led to an improvement in subsequent efforts. As a consequence, he had grown to be generous with negative feedback and had become rather wary of giving too much praise.

Kahneman immediately spotted that it was regression to the mean at work. He illustrated the misconception by a simple exercise you may want to try yourself. He drew a circle on a blackboard and then asked the officers one by one to throw a piece of chalk at the center of the circle with their backs facing the blackboard. He then repeated the experiment and recorded each officer's performance in the first and second trial.

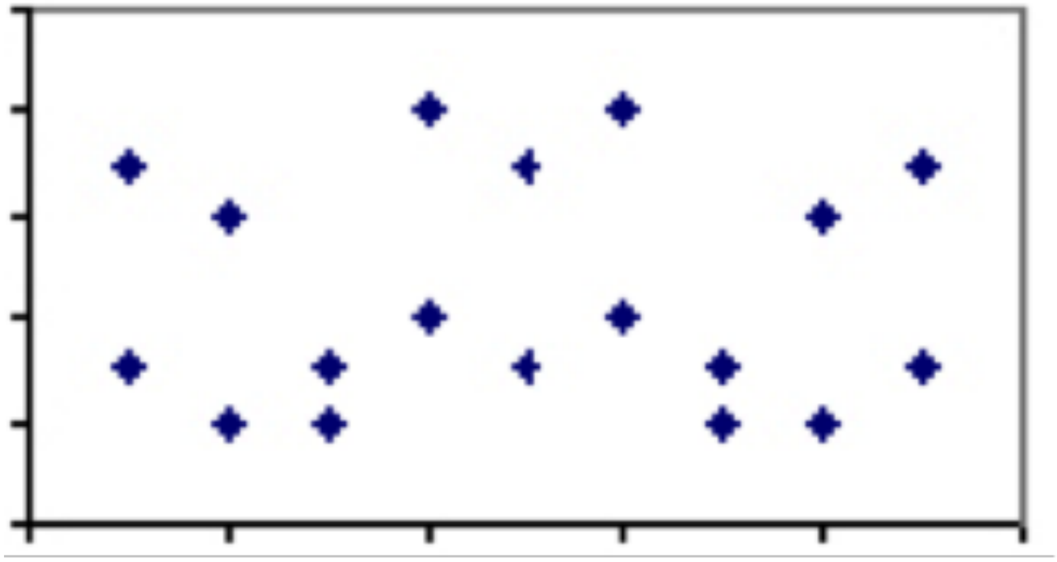
Naturally, those that did incredibly well on the first try tended to do worse on their second try and vice versa. The fallacy immediately became clear: the change in performance occurs naturally. That is not to say that feedback does not matter at all – maybe it does, but the officer had no evidence to conclude it did.

The Imperfect Correlation and Chance

At this point, you might be wondering why the regression to the mean happens and how we can make sure we are aware of it when it occurs.

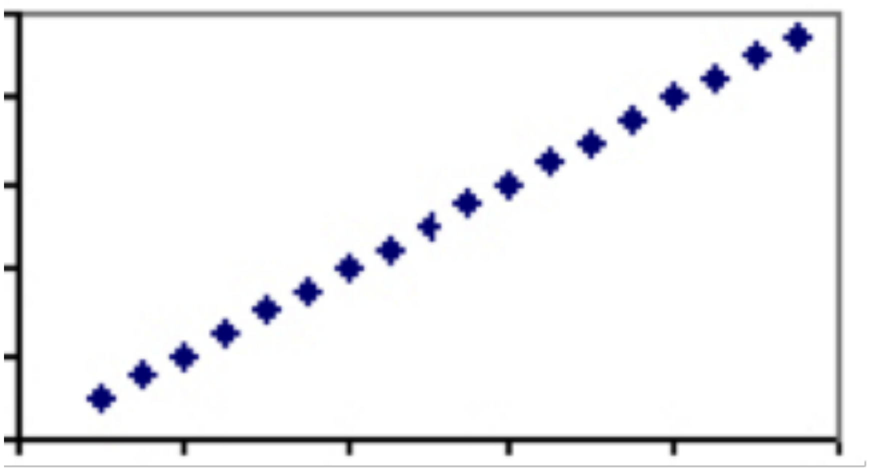
In order to understand regression to the mean, we must first understand correlation.

The correlation coefficient between two measures, which varies between -1.000 and +1.000, is a measure of their association. For example, bottled water consumption versus suicide rate, should have a correlation coefficient of close to 0. That is to say, if we looked at all countries in the world and plotted suicide rates of a specific year against per capita consumption of bottled water, the plot would show no pattern at all.



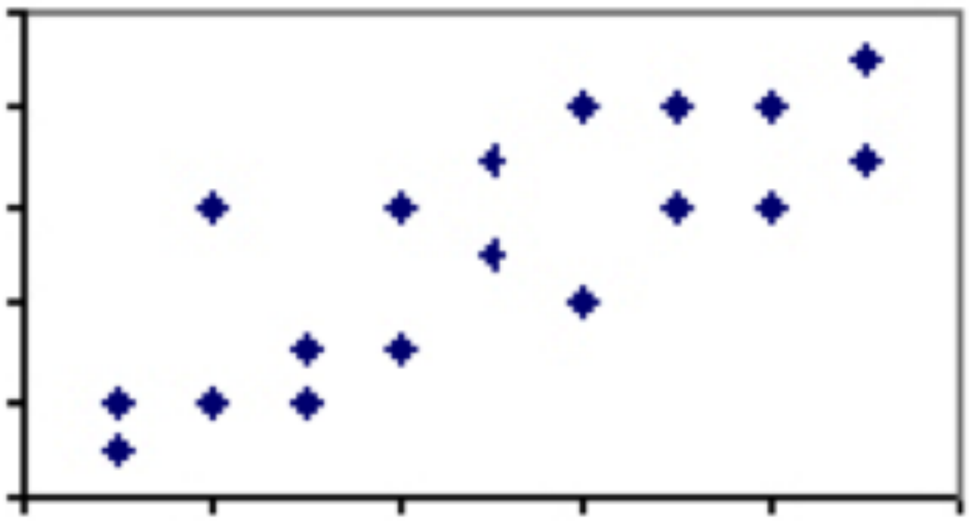
No Correlation

In contrast, there are measures that are tightly associated. A good example of this is temperature. The only factor determining temperature – velocity of molecules — is shared by all scales, hence each degree in Celsius will have exactly one corresponding value in Fahrenheit. Therefore, temperature in Celsius and Fahrenheit will have a correlation coefficient of 1.000, and the plot will be a straight line.



Perfect Correlation

Few if any phenomena in human sciences have a correlation coefficient of 1.000. However, plenty have a weak association or even a moderate association. Consider the correlation between height and weight. While virtually every three-year-old will be lighter and shorter than every grown man, not all grown men or three-year-olds who are the same height as each will weigh the same amount as each other.



Moderate Correlation

This variation and the corresponding lower degree of correlation implies that, while height is generally speaking a good predictor, there clearly are factors other than the height at play. When the correlation of two measures is less than perfect, we must watch out for the effects of regression to the mean.

The Cause, Effect, and Treatment

We should be especially wary of the regression to the mean phenomenon when trying to establish causality between two factors. Whenever correlation is imperfect, the best will always appear to get worse and the worst will appear to get better over time, regardless of any additional treatment. This is something that the general media and sometimes even trained scientists fail to recognize.

Consider the example Kahneman gives:

> *Depressed children treated with an energy drink improve significantly over a three-month period. I made up this newspaper headline, but the fact it reports is true: if you treated a group of depressed children for some time with an energy drink, they would show a clinically significant improvement. It is also the case that depressed children who spend some time standing on their head or hug a cat for twenty minutes a day will also show improvement.*

Whenever coming across such headlines it is very tempting to jump to the conclusion that energy drinks, standing on the head or hugging cats are all perfectly viable cures for depression. These cases, however, once again embody the regression to the mean:

> *Depressed children are an extreme group, they are more depressed than most other children—and extreme groups regress to the mean over time. The correlation between depression scores on successive occasions of testing is less than perfect, so there will be regression to the mean: depressed children will get somewhat better over time even if they hug no cats and drink no Red Bull.*

We often mistakenly attribute a specific policy or treatment as the cause of an effect, when the change in the extreme groups would have happened anyway. This presents a fundamental problem: How can we know if the effects are real or simply due to variability?

Luckily there is a way to tell between a real improvement and regression to the mean. That is the role of the control group who are expected to improve by regression alone. The aim of the research is to determine whether the treated group improve more than regression to the mean can explain.

Awareness of regression to the mean itself is already a great first step towards statistical thinking. If there is anything to be learned from the regression to the mean it is the importance of track records rather than relying on one-time success stories.