Regression Toward the Mean [1]

Siddharth Kalla [2]

Averaging Out Top Performances

Regression toward the mean refers to the phenomenon of ‘averaging out’ in statistics.

When a non-random sample is selected, the average of that sample tends to regress towards the mean. This is a commonly occurring phenomenon that should be taken into consideration while designing experiments.

When a non-random sample is selected, the average of that sample tends to regress towards the mean. This is a commonly occurring phenomenon that should be taken into consideration while designing experiments.

> For example, Olympic gold medalists tend to have a dip in performance after the Olympics. Regression toward the mean suggests that this is because they were the ones hitting their peak in performance (and maybe even their 'luck-peak'?) in the Olympics and therefore also gets much attention afterward, because people expect more from these athletes.
>
> The same is often true for politicians having a surprisingly good election or after the stock exchange is hitting a record level on the index.
>
> This phenomenon can be observed in a range of areas in the society, such as economy, politics, sports and even in evolution

Suppose you give a group of students a test, called pretest and select the group that lies at the bottom 5% of the total test takers. Regression toward the mean [3] implies that in the next test, the posttest, the same group will often have a higher score than the pretest values. (Pretest-Posttest-Design [4])

This means that collectively, the score of this group that initially were in the bottom 5% will no longer be in the bottom 5% and will increase in performance, regressing towards the mean.

In the similar manner, if you take the top 5% students on the pretest, they will probably perform poorer in the posttest compared to the pretest, when considered as a group.

Therefore if a program is designed that takes the less performing students and aims to improve their performance, then the efficiency of such a program should not be calculated

only on the basis of pretest and posttest scores, since an improvement will happen irrespective of such a program.

Therefore in most such studies, it is important to have a control group and compare results across these groups so that the real effect of such a program can be studied. Therefore regression toward the mean becomes an important practical phenomenon.

Understanding regression toward the mean is easy. When you select the bottom 5% of the students, it is statistically unlikely that those exact same set of students will again perform poorly on the next test as well. Even if one of the students in the group performs better, then the group as a whole is in a better position.

It is the same case with the top 5% students as well.

Therefore regression toward the mean is a statistical phenomenon that occurs in most groups. From this explanation it is also clear that the more extreme sample you select for your pretest, the higher likelihood of a regression toward the mean in the posttest.

It is important to keep in mind that regression toward the mean is a statistical phenomenon. There can be a shift of data towards the mean value due to some other parameters in the experiment after the pretest results. There can therefore be causal regressions towards the mean, which are different from the statistical phenomenon.

Links
[1] https://explorable.com/regression-to-the-mean
[2] https://explorable.com/users/siddharth
[3] http://www.socialresearchmethods.net/kb/regrmean.php
[4] https://explorable.com/pretest-posttest-designs