Re: Stationary means

Source: http://sci.tech−archive.net/Archive/sci.stat.edu/2006−09/msg00007.html

---

- *From*: Richard Ulrich <Rich.Ulrich@xxxxxxxxxxx>
- *Date*: Sun, 03 Sep 2006 21:10:51 −0400

On 3 Sep 2006 16:37:01 −0700, "Jason Foster" <retsofaj@xxxxxxxxx>
wrote:

> I don't think that this is a FAQ, but if it is I apologize for the
> noise...
>
> Discussions about "regression to the mean" tend to focus on fallacies
> relating to heights, grades, sickness, etc. Something I have not seen
> discussed is when/whether it is appropriate to assume a stationary mean
> (or, I think alternatively, a fixed distribution)?
>
> Assuming a constant distribution and repeated sampling, I can
> intuitively understand regression to the mean. However, I can imagine
> situations where the distribution is changing over time. For example
> (and here I'm talking outside of my area of expertise) the mean height
> in North America is increasing over time (ostensibly due to dietary and
> health factors). If this is the case, then what would "regression to
> the mean" mean? Towards which mean would the regression take place?

"Regression to the mean" is more general than you describe.

The X and Y variable are not necessarily the same. The
predictions based on X−deviations−from−X−mean are in
terms of Y−deviations−from−Y−means.

Thus, the child−population predictions regress to the child−mean.

> How would an observer know that a regression analysis is appropriate?
>
> Any thoughts (or pointers to resources) on the issues would be
> gratefully appreciated.

The difficult version of problem arises when

(a) groups are selected by initial "ability"; and
(b) the outcome is later "ability"; and
(c) the critic argues that initial ability will predict growth.

Thus, the critic argues that *untreated* groups will grow further apart. The naive (null) hypothesis would be that the untreated groups will tend to converge.

That illustrates the problem of covariance when the covariates are not equal at the start −− The data on hand, concerning a treatment where groups were not matched, cannot tell you which of these two circumstances was true. The case has to be argued from other data.

That was the situation in the early evaluation of the efficacy of U.S. preschool Headstart" programs for disadvantaged children. The disadvantaged children did not fall further behind. The regression model would have "expected" convergence.

Was it a realistic expectation that these children would have fallen further behind, without a program?
(I think that is the argument that won, though the technical debate may have been pushed aside by new data. I could not find much on the discussion when I looked for it a few years ago. Harvard Educational Review comes to mind.)

Hope this helps.

−−
Rich Ulrich, wpilib@xxxxxxxx
http://www.pitt.edu/~wpilib/index.html
.