Psicothema
2000. Vol. 12, nº 1, pp. 145-149

ISSN 0214 - 9915 CODEN PSOTEG

Regression toward the mean associated with extreme groups and the evaluation of improvement

Orfelio G. León and Manuel Suero
Universidad Autónoma de Madrid

Regression toward the mean effect (tendency of extreme data to resemble their mean when measurements are repeated) threatened internal validity when, in a group of extreme subjects, we attempted to test whether pre/post difference in values is due to our intervention. We are still unaware how and how much it affects: a) as a function of selection value, and b) statistical testing. This study demonstrates detection, measurement and correction of the effect. Its objective was achieved through a simulation of extreme groups, measured before and after a null treatment. The effect was measured by comparing simulated with assumed sample distribution. Correction was made by calculating the value of the alpha of the testing of the null hypothesis needed to cancel out the effect.

Regresión a la Media en la Selección de Grupos Extremos y la Valoración del Cambio. El efecto de regresión a la media (tendencia de los datos extremos a parecerse a su media cuando se repiten las mediciones) se convierte en una amenaza a la validez interna cuando, en un grupo de sujetos extremos, tratamos de probar que la diferencia entre valores pre y post se debe nuestra intervención. Hasta ahora no sabemos cómo y cuánto afecta, en función de la selección del grupo y sobre el contraste estadístico. En el presente trabajo se muestra cómo se puede detectar, medir y corregir. El objetivo se llevó a cabo mediante simulación de grupos extremos, medidos antes y después de una intervención nula. La medición del efecto se hizo comparando la distribución muestral simulada con la supuesta. La corrección se hizo calculando cuánto debía valer el alfa del contraste de la hipótesis nula para anular el efecto.

The problem of regression toward the mean as a confounded variable in the analysis of change has been included in most of Research Methods textbooks; these books guard against the threat of regression toward the mean with extreme groups in the pre-post designs but they do not state how to calculate the extent of the effect (e.g. Baker, 1994, pp. 221-22; Breakwell, Hammond and Fife-Schaw, 1995, pp. 91-92; Cook and Campbell, 1979, pp. 52-53; Haimsom and Elfenbein, 1985, p. 119; Heiman, 1995, pp.341-42; Kendall and Butcher, 1982, p.224; Shaugnessy and Zechmeister, 1997, pp. 347-348 Ware and Brewer, 1988, pp. 39-41). In this study we aim to achieve this measurement.

Some works (Gottman and Rushe, 1993, Speer 1992, 1993 and Hsu 1995) have proposed ways to achieve this measurement in order to estimate when there will be a reliable change for a person, through an appropriate correction. (e.g., double the error of measurement or of the error of prediction.) In our view, two questions remain unresolved: 1) If It has applied a treatment to an extreme group in one variable: is that correction independent of the way subjects select? 2) Should It take into account regression toward the mean when carrying out the testing of statistical hypotheses? The objective of this study is to respond to these two questions. This will be achieved by studying the phenomenon through simulation.

Extreme groups. Let us suppose that we select people to form part of a treatment group due to their having obtained extreme values in a test. The majority of those selected will be genuine extreme value subjects, but some will be included because the errors of measurement were high. In the second measurement they will return to their true, lower values. This deviation is confounded with the effect of the intervention. Obviously, the higher, the value of the selection the greater will be the errors. Those writing in Research Methods.

Testing of statistical hypotheses. Let us suppose that we have carried out the psychological intervention and that we have obtained the pre and post measures. We apply a difference of means test, with repeated measures. If regression toward the mean has introduced an effect that adds to the effect of the treatment, we will have to take it into account in the analysis. Should we continue to assume that the difference of the means of the null hypothesis is zero? We propose that it is necessary to reanalyze the data, assigning to the difference of means of the hypothesis the value of the regression effect- in our case, obtained in the simulation. That is, if the intervention is effective the mean of the post-test scores would be greater than the pre-test plus the effect of regression.

Finally, the simulation will create thousands of situations like those described. We shall measure the pre and post-test scores, the extent of the effect of regression and the appropriate way to correct the statistical hypothesis test.

Correspondencia: Orfelio G. León
Facultad de Psicología
Universidad Autónoma de Madrid
28049 Madrid (Spain)
E-mail: orfelio.leon@uam.es

Method

In order to evaluate the effect of regression toward the mean we carried out a simulation of a pre-post design, with a single group, in which the treatment was null.

Basically, the simulation followed these steps: 1) First measurement (pre): A group was created in which the subjects belonging to it had a score equal to or greater than a selection value in a fixed test. In this way, the first measurement defined an extreme group; 2) Second measurement (post): Using the same test, for each one of the subjects in the group defined in the previous step, a second measurement was taken. The differences between the pre and post measurements were then calculated.

Both in the simulation phase and in the analysis of the differences found we have followed the assumptions of the classical theory of tests (Gulliksen, 1950; Lord and Novick, 1968). We present these assumptions below, with the object of explaining how they affect the differences between the measures, and how they determine the way in which the simulation is carried out.

1. The test is not perfect: errors are made in the measurement. The test produces an empirical score (X), which can be broken down into a true score (V) and an added error (E) associated with the measurement.

2. Given that the treatment between the two measurements is null, the true score (V) is maintained constant for each subject and for each measurement.

3. The true score (V), error of measurement (E) and empirical score (X) are random variables with known distribution and parameters.

4. The errors of measurement (E) are independent of the true score (V).

5. The errors of measurement associated with the second measurement are independent of the errors associated with the first.

6. The error is a normally-distributed random variable with parameters $\mu_E=0$ and σ_E.

From these assumptions it can be concluded that the difference obtained between the pre and post depends exclusively on the errors. In the pre-test measurement the score obtained for the subject ith is equal to:

$$(1) \quad x_{i1}=v_{i1}+e_{i1,}$$

whilst the post score obtained is:

$$(2) \quad x_{i2}=v_{i2}+e_{i2}$$

The difference for the subject ith is:

$$(3) \quad d_i=x_{i1}-x_{i2}$$

In accordance with the second assumption and equations (1) and (2), equation (3) is expressed as:

$$(4) \quad d_i=e_{i1}-e_{i2}$$

Thus, the difference found for the subject ith depends exclusively on the errors of measurement. There will be a regression toward the mean effect if the difference is positive. In order to quantify this effect we need to know the errors committed in the two measurements.

In an empirical work it is not possible to know with precision the errors committed, but they can be known exactly in a simulation process. Thus, the simulation of the pre-test measurement was carried out in the following way:

1. In accordance with the third assumption, we obtained randomly a true score v_{i1}.

2. In accordance with the sixth assumption, we obtained randomly an error of measurement e_{i1}.

3. If the sum $v_{i1}+e_{i1}$ is superior to a selection value the scores v_{i1} and e_{i1} are accepted. If not, new values of V and E are taken.

For the post-test measurement, and in accordance with the second assumption, it was only necessary to randomly take values of the error. We thus obtained e_{i2} .

At the end of the simulation we obtained, for each subject belonging to an extreme group, three scores:1) True score, 2) Error committed in the pre-test measurement, and 3) Error committed in the post-test measurement. The difference, for each subject, between the empirical pre and post measurements was equal to the difference $e_{i1}-e_{i2}$; this magnitude indicates the effect of regression toward the mean (K_R).

In accordance with what we have presented up to now, for carrying out the simulation it is necessary to know: 1) $f_V(v)$: Function of the density of probability of the true scores, 2) $f_E(e)$: Function of the density of probability of the errors, and 3) Selection value for defining the extreme group. In addition, it is necessary to determine the size of the group (n) and the number of groups (N).

In the study described here three tests were used, WAIS, MMPI (depression), EPI (extroversion, males). The $f_V(v)$ and $f_E(e)$ were obtained from the test manuals. For each test we carried out a simulation with an n equal to 30 and an N equal to 10,000, and three selection values that defined the extreme group. Table 1 shows the values of μ and σ for each $f_X(x)$, $f_V(v)$ and $f_E(e)$. The selection values (expressed in standard and empirical scores) appear in Table 2.

The simulation was carried out using a PC, by means of a program in C language designed by the authors.

Results

At the end of each simulation (test x selection value) means and standard deviations of empirical scores for the conditions pre, post and difference pre-post were obtained. This data is shown in Table 2.

Analysis of the results for correcting the effect of regression toward the mean. This section can be more easily followed by referring to Figure 1.

We call regression sampling distribution/H_o (H_0:$\mu_{pre}-\mu_{post}=K_R$) that which is obtained with the simulation and standard sampling distribution/H_o (H_0:$\mu_{pre}-\mu_{post}=0$), that assumes that the dif-

Table 1
Parameters of the Simulation

Variables	Test		
	WAIS	MMPI (depression)	EPI (extroversion, M.)
μ_X	100	21.36	10.72
σ_X	15	3.99	3.96
σ_V	14.621	4.61	3.245
σ_E	3.35	2.31	2.27

Table 2
μ y s of the Simulated Sampling Distributions of the Means

Selection value		Test: WAIS	Test: MMPI	Test: EPI
z=.67		X_{min}=110.05	X_{min}=24.45	X_{min}=13.37
	Pre	μ =119.01, σ =1.37	μ = 27.20, σ =.41	μ =15.74, σ = .36
	Post	μ =118.07, σ =1.54	μ =25.74, σ =.63	μ =14.08, σ =.58
	Difference	μ = .94, σ = .86	μ = 1.46, σ = .57	μ = 1.66, σ = .55
z=1.28		X_{min}=119.28	X_{min}=27.28	X_{min}=15.81
	Pre	μ =126.35, σ =1.11	μ = 29.46, σ =.35	μ =17.68, σ =.30
	Post	μ =125.04, σ =1.36	μ =27.43, σ =.62	μ =15.41, σ =.58
	Difference	μ = 1.31, σ = .86	μ = 2.03, σ = .56	μ =2.27, σ = .55
z=1.67		X_{min}=125	X_{min}=28.17	X_{min}=16.26
	Pre	μ =131.22, σ =1.01	μ =30.21, σ =.33	μ =18.06, σ =.29
	Post	μ =129.67, σ =1.29	μ =27.99, σ =.61	μ =15.64, σ =.57
	Difference	μ = 1.55, σ = .86	μ = 2.22, σ = .56	μ = 2.42, σ = .55
z=2.05		X_{min}=130.00	X_{min}=29.54	X_{min}=17.37
	Pre	μ =135.60, σ =.92	μ =31.39, σ =.30	μ =19.01, σ =.27
	Post	μ =133.82, σ =1.21	μ =28.88, σ =.60	μ =16.29, σ =.57
	Difference	μ = 1.78, σ = .86	μ = 2.51, σ = .56	μ = 2.72, σ = .55

Note: Difference = Pre - Post

ference of populational means is zero. The analysis consists in making two calculations: 1) Using K_R, the calculation of the probability of rejecting the null hypothesis of no difference when in fact it is correct (type 1 error); 2) Using K_R, the calculation of the alpha we would have to use under H_0:μ_{pre}-μ_{post}= 0 in order to maintain constant the probability of committing a type 1 error. We shall continue by demonstrating how these calculations were made for one case:

DATA: Test= WAIS; selection value z= .67; one-tailed alpha level of .05, z= 1.64; H_0:μ_{pre}-μ_{post}= 0; H_0:μ_{pre}-μ_{post}= K_R= .94; σ= .86 (see Table 2).

1. Probability of committing a type 1 error

1.a. (D˜0)/.86= 1.64; D˜= 1.4104; this implies that, under the usual hypothesis H_0:μ_{pre}-μ_{post}=0, if we find a pre-post difference greater than 1.4104 we reject the zero difference, concluding that the intervention has been significant.

1.b. p(z ≥ ((1.4104-.94)/.86)= .2922; this implies that, under the hypothesis H_0:μ_{pre}-μ_{post}=K_R, the probability of finding pre-post differences greater than 1.4101 is in fact .2922 (not .05). In order to calculate the probability of type 1 error, it is necessary to subtract .05, given that the values of the distribution/H_0:μ_{pre}-μ_{post}=K_R in the right tail, with a probability equal to or greater than .05, are also rejected under H_0:μ_{pre}-μ_{post}=0.

2. Alpha corrected

2.a. (D˜.94)/.86 = 1.64; D˜= 2.3504; this implies that, under the hypothesis H_0:μ_{pre}-μ_{post}=K_R, if we find a pre-post difference greater than 2.3504 we reject the zero difference, concluding that the intervention has been significant.

2.b. p(z ≥ ((2.3504-0)/.86)= .003138; this implies that, under the hypothesis H_0:μ_{pre}-μ_{post}=0- that which is used by the researcher-, we must correct the alpha to .003138 so that the rejection value is 2.3504. This allows us to maintain the probability of type 1 error at .05.

These two calculations (1 and 2), (see Table 3), show the probability of type 1 error and the necessary correction, for the tests studied and the selection values employed.

Correction of regression effect for other tests. We studied this problem for the selection value z=.67, given that it is more frequent and that more extreme values involve greater difficulties for researchers. With the data obtained we have constructed Figure 2, which relates the corrected alpha with S^2e/S^2x of the test. We have used as asymptote the case of S^2e/S^2x=0, in which the corrected alpha coincides with the initial alpha (.05).

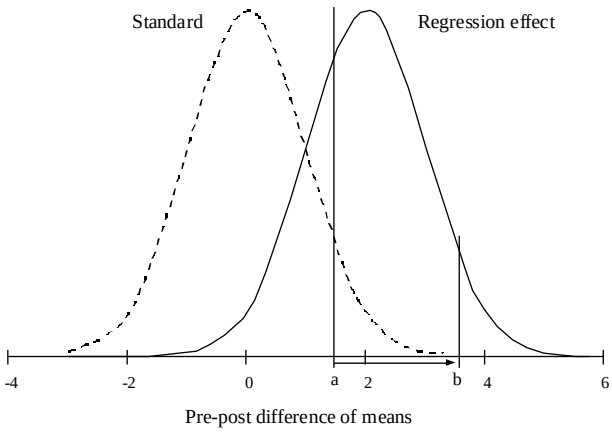

Figure 1: The standard model is that used to check H_0:μ_{pre}-μ_{post}=0. The regression model is obtained by simulation for a specific case: H_0:μ_{pre}-μ_{post}=2. When we obtain pre-post values between (a) and (b) we reject H_0 with the standard model. The correct conclusion would be to accept it, given that there is a regression effect that has displaced the distribution

Table 3
Measure of the Regression Toward the Mean Effect Asociated to the Selection Value

Test	Z	%	p(type 1 error)	Corrected Alpha
WAIS	0.67	25	0.2422	0.00314
	1.28	10	0.4035	0.00078
μ=100	1.67	4.75	0.5145	0.00029
σ=15; σE=3.35	2.05	2.28	0.6163	0.00010
MMPI	0.67	25	0.7716	$1.33\ 10^{-5}$
Depression	1.28	10	0.9264	$7.01\ 10^{-8}$
μ=21	1.67	4.75	0.9399	$1.05\ 10^{-8}$
σ=4.61; σE=2.31	2.05	2.28	0.9478	$4.62\ 10^{-10}$
EPI	0.67	25	0.9049	$1.60\ 10^{-6}$
Extroversion, M.	1.28	10	0.9918	$4.04\ 10^{-9}$
μ=10.72	1.67	4.75	0.9962	$7.71\ 10^{-10}$
σ=3.96; σE=2.27	2.05	2.28	0.9993	$2.27\ 10^{-11}$

If we consider that the reliability of other tests may be between the values studied (.95 to .67), the general graph of Figure 2 makes interpolation difficult. With the object of showing in more detail the interpolations between the values studied we have constructed Figure 3 for tests with S^2e/S^2x between 0.05 and 0.33.

Discussion

Concerning the first objective of this study: the independence of possible corrections in the post scores with respect to the selection point of the subjects. The correction would be independent if K_R were constant for all the different selection points. As can be observed from the data in Table 2, the K_R (difference pre-post) increases as the selection value increases (e.g. for the MMPI, from 1.46 to 2.51).

Concerning the second objective of the study: consequences of the K_R in the testing of statistical hypotheses. These consequences have been studied in terms of the modification of the type 1 error committed in the testing. As can be observed from the data in Table 3, extensive modifications are produced which depend on the reliability of the test and on the selection point (e.g. for the EPI/z=.067 p(type 1 error)=.90).

Concerning the third objective: correction of the regression effect. The effect is corrected by modifying the alpha in the testing of hypotheses (Table 3). To be able to make this modification it is necessary to know K_R, which we have ascertained by means of the simulation. For other tests, we have constructed Figures 1 and 2. With the value of S^2e/S^2x of the test, we can make an approximate estimation of the value of alpha we must use in order to maintain constant the probability of type 1 error.

Other conclusions: a) the data obtained supports the proposals of Speer (1993) and Hsu (1995) that the magnitude of the correction (K_R) should be inversely proportional to the reliability of the test; however, the values proposed are greatly superior to those found in the simulation, so that these proposals are quite conservative and difficult to fulfill. We found for the MMPI-selection value z=2.05, α=.05-a critical difference/ H_0:μ_{pre}-μ_{post}=0 of D˜=3.43; on the other hand, they would propose as critical difference double the standard error of measurement, 2x4.61=9.22. b) The combination of low reliability and a high cut-off point is dramatically manifested in a regression effect which causes the post-test empirical mean to be inferior to the selection value (e.g. EPI/z=2.05; X_{min}=17.37; μ_{post}=16.29) (see Table 2).

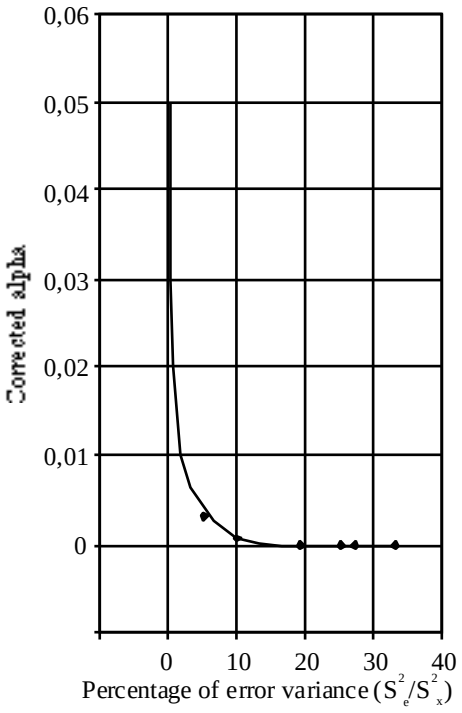

Figure 2: Variation of the corrected alpha in function of the error of measurement of the test

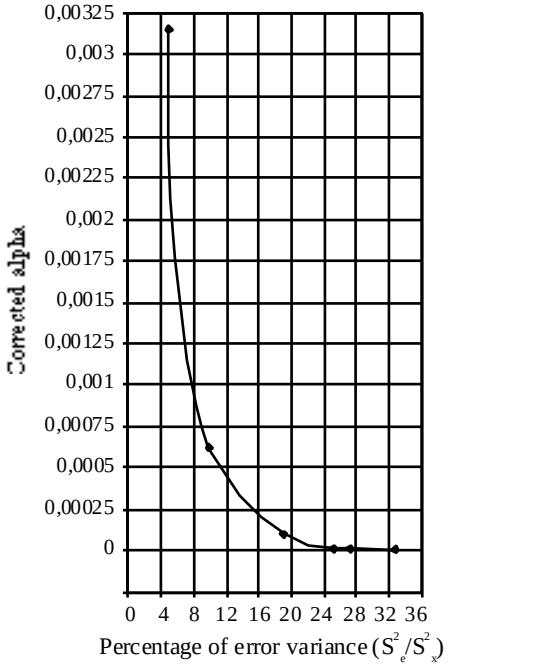

Figure 3: Figure for estimating the corrected alpha. Situate the value S^2e/S^2x of the test on the x-axis. Plot a perpendicular cutting the curve. Plot a horizontal cutting the y-axis. Interpolate. (Calculations for a selected value z=.67 and for tests with rxx between .95 and .67)

Final note: the simulation procedure has shown itself to be of great utility for demonstrating the regression toward the mean effect and its measurement. In research using the pre-post design with extreme groups, the failure to correct this effect will involve a high probability of concluding that a psychological intervention has been significant, when in fact this is not the case. For a general solution to the correction of the regression effect it is necessary to construct graphs such as that presented here for other selection values, or to find an analytical solution that expresses the value of K_R.

Referencias

Baker, T. (1994). *Doing social research (2ed.).* New York: McGraw-Hill.

Breakwell, G.M., Hammond, S. &, Fife-Schaw, C. (Eds.) (1995). *Research methods in psychology*. London: Sage.

Cook, T.D. &, Campbell, D. T. (1979). *Quasi-experimentation. Design and analysis issues for field settings*. Boston: Houghton Mifflin Co.

Gottman, J.M.&, Rushe, R.H. (1993). The analysis of change: issues, fallacies, and new ideas. . *Journal of Consulting and Clinical Psychology, 61*, 907-910.

Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley & Sons.

Haimsom, B.R. &, Elfenbein, M.H. (1985). *Experimental methods in psychology*. New York: McGraw-Hill.

Heiman, G.A. (1995). *Research methods in psychology*. Boston: Houghton Mifflin Co.

Hsu, L.M. (1995).Regression toward the mean associated with measurement error and the identification of improvement and deterioration in psychotherapy. *Journal of Consulting and Clinical Psychology, 63*, 141-144.

Kendall, P.C. &, Butcher, J.N. (Eds.)(1982). *Handbook of research methods in clinical psychology*. New York: John Wiley & Sons.

Lord, F.M., & Novick, M.R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Shaughnessy, J.J., & Zechmeister, E.B. (1997). *Research methods in psychology* (4 ed.). New York: McGraw-Hill.

Speer, D.C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60*, 402-408.

Speer, D.C. (1993). Correction to Speer. *Journal of Consulting and Clinical Psychology, 61*, 27.

Ware, M.E. &, Brewer, C. L.(Eds.) (1988). *Handbook for teaching statistics and research methods*. Hillsdale, NJ: LEA.

Aceptado el 16 de marzo de 1999