Review of Psychology,
2009, Vol. 16, No. 2, 65-76 UDC 159.9

Comparison of the power of four statistics in repeated measures design in the absence of sphericity with and without serial autocorrelation

PAULA FERNÁNDEZ, GUILLERMO VALLEJO, PABLO LIVACIC-ROJAS,
JAVIER HERRERO and MARCELINO CUESTA

The present article examines the behaviour of four univariate statistics for analyzing data in a mixed repeated measures design, the procedures of Greenhouse and Geisser (1959), of Lecoutre (1991), of Hearne, Clark and Hatch (1983) and of Jones (1985), which differ in how they approach the absence of sphericity, assuming either arbitrary correlation or serial autocorrelation. These four approaches were compared with respect to empirical power in conditions of multivariate normality and absence of normality, and of different underlying structures of covariance. Overall, when the distribution is normal, Monte Carlo comparisons indicate that when the matrix is stationary autoregressive or structured non-stationary autoregressive, the Lecoutre and Hearne et al. statistics are more powerful, the former enjoying slightly higher empirical power, with no large differences between the two in either direction of the autocorrelation (positive and negative first-order serial correlation). For an arbitrary non-stationary matrix, the Hearne et al. procedure is considerably more powerful than the Lecoutre statistic when the deviation of the sphericity is slight and severe, both in the two directions of the autocorrelation (positive and negative first-order serial correlation) and when it is arbitrary (correlation=0). When the data are underlain by a non-normal distribution, the HCH procedure is that with the greatest empirical power when the serial correlation is positive, and the JN procedure when the serial correlation is negative whatever the underlying deviation matrix.

*Key words***:** Power of the test, stationary autoregressive matrix, structured non-stationary autoregressive matrix, arbitrary non-stationary autoregressive matrix, non-stationary matrix with arbitrary correlation.

The repeated measures designs provide an efficient strategy for examining the effect of treatments administered consecutively or examining the evolution of behaviour over time, given that they permit the extraction of individual differences from the experimental error. It is for this reason that they have become the most widely used tool for diagnosing, explaining and predicting biological, psychological and social processes (Keselman, Algina, & Kowalchuk, 2001). The most common type of design involves two factors, between-subjects (A) and within-subject (B). Subjects (i=1,..., n_j, $\Sigma\ n_j=N$), classified according to, or assigned at random to the between-subjects levels (j=1,..., p), are observed and measured at all levels of the within-subject factor (k=1,..., q), be they different treatments or a small number of measurement occasions that result from the systematic choice of fixed and equidistant time intervals.

Given that the present work and the recommendations made at the end of it are addressed to psychologists carrying out applied research, with a view to achieving greater contextualization of the application of the statistics tested here, we begin by describing two real studies:

Study 1 (S.1): Palmero, Breva, Diago, Díez and García (2002) carry out an experiment for exploring the role of Type A and Type B emotional patterns (between-subjects classification variable: n_A=28 and n_B=28) in activation, variability and cardiac recovery (dependent variables), in three situations represented by the within-subject independent variable (habituation, task –presentation of stressful stimuli– and recovery).

Paula Fernández, University of Oviedo, Departamento de Psicología, Plaza Feijóo, s/n, E-33003 Oviedo, Spain, E-mail: paula@uniovi.es (the address for correspondence)

Guillermo Vallejo, University of Oviedo;

Pablo Livacic-Rojas, University of Santiago de Chile;

Javier Herrero, University of Oviedo;

Marcelino Cuesta, University of Oviedo.

Acknowledgements

This study was carried out with the help of grant MSI-PSI2008-03624, from the Spanish Ministry of Science and Innovation.

Study 2 (S.2): Méndez, Orgiles and Espada (2004) carry out another experiment to test the effectiveness of the staged emotional situations program for the treatment of nyctophobia (fear of the dark). Forty-five children were assigned at random to two experimental conditions: treatment group (staged emotional situations) or control group (waiting list). This was the between-subjects independent variable. The two groups were observed at four time points –pretest, postest, 3-month follow-up and 6-month follow-up (assigned within-subject independent variable).

The nature of the *q* levels of the within-subject variable (S.1=3 levels; S.2=4 levels) will determine not only the nature of the dependent variable, but also certain characteristics that may accompany it. The dependent variable may be the same in each measurement made for each one of the treatments administered randomly for each subject, thereby avoiding order effects; moreover, if a prudent distance is kept between applications, residual effects will also be avoided. In this case, the variance-covariance matrix underlying the data may be spherical (this is the expectation in S.1). Sphericity is present when the variances corresponding to the differences between the different measurement occasions are equal, or when there are equal variances and equal covariances. The matrices represented by the first condition are called Type H or Huynh-Feldt (1970). Those represented by the second condition are called Type S or Combined Symmetry, and are a particular case of Type H matrices. In both matrices there is a correlation between subjects' responses, but this correlation is constant.

It may be essential for the dependent variable to be recorded on different measurement scales for each treatment. Another possibility is for it to be the same in each application, but with the measurement done in accordance with age or time, rather than with the different treatments. In the latter two situations the variance-covariance matrix most probably deviates from sphericity, though for quite different reasons. In the first case, the variances of the treatments might simply be arbitrarily heterogeneous (leading at the same time to a heterogeneity of covariances, also without a defined structure), so that the correlation matrix is arbitrary. In the second case, it is very probable that we will observe certain trends related to maturation or learning processes producing residual and/or autocorrelation effects and, eventually, giving rise to a certain heterogeneity in the variances of the treatments (this is the expectation in S.2).

The importance of the above is crucial for deciding on the statistical technique to use for testing the different null hypotheses of the design (see Fernández, Livacic-Rojas, & Vallejo, 2007 for a full review of them, both univariate and multivariate). It is well known that the power of multivariate techniques is low when samples sizes are relatively small; moreover, if the number of subjects in each group is smaller than the number of repeated measures minus one (it should hold that $N\text{-}p \geq q\text{-}1$), no multivariate statistic can be calculated because the covariance matrix will be singular. In this case the researcher will only be able to use univariate statistics. Univariate Analysis of Variance (AVAR) is undoubtedly the most powerful technique when the assumptions of the model are satisfied (Keselman, Lix, & Keselman, 1996), that is, multivariate normality, independence between the vectors of observations of the different experimental units, homogeneity of the dispersion matrices (Σ), and their sphericity. The immediate consequence of violating the assumption of sphericity is that the null hypotheses of the effects of the design are falsely rejected more often than they should be, and more so the greater the deviation (Collier, Baker, Mandeville, & Hayes, 1967). In order to solve this problem, and depending on whether or not the remaining assumptions of the AVAR are met, different authors have developed different univariate alternatives, examples being the Greenhouse and Geisser (1959) ($\hat{\varepsilon}$), Huynh and Feldt (1976) ($\tilde{\varepsilon}$) and Lecoutre (1991) ($\tilde{\varepsilon}_L$) tests, (Quintana & Maxwell, 1994; Fernández et al., 2007 have carried out thorough reviews). All of these aim to correct the critical values of the univariate *F* by multiplying the degrees of freedom (*df*) by a value of ε that indicates the deviation from the covariance matrix of the required sphericity pattern and which is calculated from $\hat{\Sigma}$ (covariance matrix averaged from the design data). However, all assume arbitrary correlation between scores, that is, that the correlation between observations at different time points is not a function of the time distance between them. But, as there may be serial dependence underlying the data, some authors (Hearne, Clark, & Hatch, 1983; Jones, 1985, among others) have proposed univariate models of variance that take it into account. Thus, for example, Hearne et al. (1983) and Jones (1985) give priority to first-order serial autocorrelation. The former calculates the value of $\hat{\varepsilon}$ over $\hat{\Sigma}_p$ (estimated covariance matrix assuming that the data show first-order serial autocorrelation), and the latter proposes modifying the calculation of the summed squares of the AVAR incorporating the serial correlation in them and subtracting one *df* in the within-subject error (see also, Fernández, 1995).

Another, more flexible approach to the analysis of repeated measurements, and particularly useful when sample size is sufficiently large to support asymptotic inference, is the mixed linear model (MLM). Under this approach, implemented in commercial software packages, including the widely used SAS® and SPSS programs, researchers, rather than presuming a certain type of covariance structure, may model the structure before testing for treatment effects. For example, the best covariance structure can be selected based on Akaike's Information Criterion (AIC) and/or Schwarz's Bayesian Information Criterion (BIC) values for various potential covariance structures. According to advocates of the mixed-model approach, selecting the most parsimonious covariance structure possible is very important, as it may result in more accurate and efficient inferences of the fixed-effects parameters of the model, and consequently more powerful tests of the treatment effects. However, it has weak points as well, and two in particular: on the one hand, potential problems with identifying the structure of the underlying Σ matrix, given that the AIC and BIC criteria do not always select it correctly (Keselman, Algina, Kowalchuck, & Wolfinger, 1999a, 1999b; Livacic, 2005; Vallejo, Fernández, & Ato, 2003); and on the other, the estimators of accuracy and inference are based on its/their asymptotic distribution –with very large samples it fits well– so that serious problems can occur when working with small samples (Vallejo et al., 2002).

There is an abundance of research testing the behaviour of univariate procedures that correct the absence of sphericity without assuming the existence of serial autocorrelation, in addition to those testing the MLM (excellent reviews can be found in Keselman et al., 1996; Keselman et al., 2001; Blanca Mena, 2004; Fernández et al., 2007). However, there are very few studies on univariate procedures that correct the absence of sphericity assuming the existence of serial correlation, so that we shall consider just two.

Recently, Fernández, Vallejo, Livacic-Rojas, Herrero and Cuesta (2008) examined the behaviour of six univariate estimators with respect to Type I error, namely, AVAR, the Greenhouse-Geisser Test (1959) (GG), the Huynh-Feldt Test (1976) (HF), the Lecoutre Test (1991) (LEC), the Hearne et al. test, (1983) (HCH), and the Jones Procedure (1985) (JN), in a split-plot factorial design ($p \times q$). Their results showed that no statistic was robust in all the conditions studied. The AVAR, HF and JN procedures displayed the poorest behaviour, since they were the most dependent of all the variables manipulated, and GG, LEC and HCH more often maintained the Type I error within the limits of robustness. The differences between them depended on the structure of the underlying covariance matrix. When the underlying matrix was stationary autoregressive and decreasing structured non-stationary autoregressive, the HCH procedure was the most robust of them all, its estimation did not depend on ρ at all (either its magnitude or its direction) or on n_j, or on q; however, when the matrix was increasing structured non-stationary autoregressive and arbitrary non-stationary autoregressive the most robust procedures were GG and LEC. When the matrix was non-structured (absence of serial correlation and of sphericity), GG displayed the best behaviour.

Subsequently, Fernández, Vallejo and Livacic-Rojas (2008) compared, by means of simulation, the GG, LEC and HCH procedures, together with the MLM (with the matrix correctly identified) under the same matricial conditions as in the previous research. The results obtained with the GG, LEC HCH procedures replicated those of the previous work. The MLM showed less robustness than the previous procedures when the underlying matrix was decreasing structured non-stationary autoregressive and when the matrix was unstructured.

In the two previous studies it was shown that the four procedures (GG, LEC, HCH and MLM) made better estimation under serial autocorrelation (better when it is positive than when it negative) than under arbitrary correlation. However, only HCH and MLM depended significantly on its magnitude, HCH making better estimation the greater the autocorrelation, and MLM doing so the smaller the correlation. GG, LEC and HCH showed better estimation the closer the matrix to sphericity, but MLM was not affected by the value of ε. For the same sample size, the GG, LEC and MLM procedures were more robust the greater the q. The HCH procedure was the least dependent on q, except when the underlying matrix was arbitrary autoregressive, in which case, for the same sample size, its estimation improved the greater the q. All the procedures improved their estimation the greater the sample size, GG and MLM being the most dependent on this variable.

This being the case, and assuming that the choice of contrast statistic for analyzing a repeated-measures design data is a univariate one, the general purpose of the present study is to evaluate the performance of the procedures which in the two previous studies were the most robust (GG, LEC, HCH) and also the procedure JN. Up to now, no research has examined the sensitivity for detecting the treatment effect of the HCH and JN procedures. The research was carried out in the same conditions as in our previous work, that is, when there is absence of sphericity in both situations, under serial correlation and under arbitrary correlation, and for data collected in the format of a split-plot factorial design ($p \times q$). As it is known that educational and behavioural research data rarely follow a normal distribution (Micceri, 1989), the procedures will be examined both under normal distribution and in the absence of normality.

METHOD

To evaluate the sensitivity of the GG, LEC, HCH and JN approaches, we carried out simulation studies for a balanced split-plot factorial design ($3 \times q$) underlain by an additive model. Bearing in mind that we have just briefly defined the three procedures, we shall not describe their formulation for two reasons: first, because the GG and LEC procedures are described in numerous publications (see, e.g., Vallejo, 1991; Fernández et al., 2007) and the HCH and JN procedures are clearly explained in Hearne et al. (1983) and Jones (1985), respectively, and secondly, because their calculation is extremely simple.

In the first study, we compared the power of the approaches proposed with data generated from multivariate normal distributions when the assumption of multisample sphericity was unfulfilled. For this purpose, three variables were manipulated: (a) sample sizes (n_j), (b) measurement occasions (q) and (c) structure of the population covariance matrix (Σ). The behaviour of the test statistics was investigated with three total sample size conditions: N=15 (n_j=5), N= 30 (n_j=10) and N= 46 (n_j=16). These sample sizes were selected because the last two (N= 30 and N= 46) are typical of what is encountered in practice, particularly in areas such as animal psychology and applied behaviour analysis, while in applied clinical psychology the sample size is frequently very small, which is why we chose N=15. There were four levels of the within-subject factor: 4, 6, 8 and 12. It is common in simulation research to study within-subject factors with 4 levels, and to a lesser extent with 8 levels. With the aim of observing the trend of the power of the test statistics depending on the number of levels, we decided to use the numbers mentioned.

Simulated data were generated using four covariance structures: stationary autoregressive (AR), decreasing and increasing structured non-stationary autoregressive (ARSH-D and ARSH-I, respectively), arbitrary non-stationary autoregressive (ARAH) and unstructured (UN). The AR matrix displays stationarity in the variances ($\sigma^2 = 10$ in our case) and the correlation between the *kth* and *k'th* observation is $\rho^{|k-k'|}$ (ρ= [-0.8:0.8: (0.2)] in our case). The ARSH matrices express covariance matrices with the same positive and negative serial correlation design as the AR matrices, but exhibit structured within-subject heterogeneity, so that the variances vary through *q* in increasing (ARSH-I) or decreasing (ARSH-D) arithmetic progression. All of these matrices have a deviation from sphericity that can be calculated by means of ε or by means of ε_p, and whose size depends on each of the elements used in their construction (Edwards, 1991). Of paramount interest for us was to observe the behaviour with respect to the power of the test of the procedures referred to above when for the same degree and intensity of autocorrelation there is a different deviation from sphericity. This aspect has not been studied up to now. In this way we could observe whether the power of a statistic was more strongly affected by one aspect or by another. For this purpose we used matrices that we called ARAH, which express covariance matrices with the same positive

Table 1

Theoretical power for the within-subjects main effect (α=.05). AR and ARSH-I (*q*=4 and 6); ARAH (ε=.50 y ε=.75) and NE (ε=.56 y ε=.75), both (*q*=4). Mean value of ε depending on Σ for GG, LEC and HCH tests

	Σ	AR		ARSH-I		ARHA ε=.50		ARHA ε=.75	
P	n_j	GG	LEC	GG	LEC	GG	LEC	GG	LEC
.20	5	.78(-)	.79(-)	.77(-)	.79(-)	.73	.74	.76	.77
	10	.82(-)	.85(-)	-(-)	-(-)	.89	.89	-	-
.40	5	.78(-)	.79(-)	.77(-)	.79(-)	.74	.75	.77	.79
	10	- (-)	- (-)	- (-)	- (-)	.89	.89	-	-
.60	5	.77(-)	.79(-)	.77(-)	.79(-)	.76	.77	.80	.80
.80	5	.77(-)	.78(-)	.76(-)	.77(-)	.78	.78	-	(-)
Pw.R		.77		.77		.75		.77	
-.20	5	.76(-)	.78(-)	.76(-)	.78(-)	.72	.73	.75	.77
-.40	5	.74(-)	.77(-)	.73(-)	.76(-)	.70	.72	.74	.76
	10	.89(-)	.89(-)	.79(-)	.79(-)	.89	.89	.88	.89
	5	.67(.79)	.69(.79)	.66(.79)	.69(.79)	.68	.70	.73	.75
-.60	10	.78(.)	.78(-)	.78(-)	.78(-)	.87	.88	.88	.88
	16	- (-)	- (-)	.79(-)	.79(-)	.65	.69	.74	.73
	5	.55(.69)	.56(.71)	.54(.69)	.56(.71)	.87	.88	.88	.89
-.80	10	.71(.78)	.71(.79)	.70(.78)	.71(.79)	.95	.97	(-)	(-)
	16	.76(-)	.75(-)	.75(-)	.76(-)				
Pw.R		**.77**		**.77**		**.74**		**.77**	
	n_j	NE ε=.56			n_j	NE ε=.75			
		GG	LEC			GG	LEC		
	5	.64	.66		5	.73	.74		
	10	.87	.87		10	.84	.86		
	16	.96	.93		16	-	-		
Pw.R		**.64**				**.73**			

Mean value of ε depending on Σ

	AR			ARSH-I			ARSH-D		
ρ	GG	LEC	HCH	GG	LEC	HCH	GG	LEC	HCH
.20	.78	.99	.95	.73	.91	.95	.78	.99	.95
.40	.75	.94	.90	.71	.88	.90	.75	.94	.90
.60	.71	.87	.82	.67	.81	.82	.71	.87	.82
.80	.65	.79	.73	.61	.72	.74	.65	.78	.73
-.20	.78	.99	.94	.72	.89	.94	.78	.98	.94
-.40	.72	.90	.84	.67	.81	.84	.72	.89	.84
-.60	.62	.74	.68	.58	.68	.70	.62	.74	.68
-.80	.48	.54	.50	.46	.51	.52	.48	.53	.50

Note. Σ= Population covariance structure; AR= Stationary autoregressive; ARSH-I Increasing structured non-stationary autoregressive; ARSH-D Decreasing structured non-stationary autoregressive; ARAH= Arbitrary non-stationary autoregressive; GG= Greenhouse-Geisser Test (1959); LEC= Lecoutre Test (1991); HCH= Hearne et al. Test (1983; n_j= subjects in each one of the groups of the between-subjects variable; ρ= autocorrelation; ε= deviation from sphericity. Under the structures of Matrix AR and Matrix ARSH-I, figures outside the brackets indicate theoretical power for *q*=4, and those inside the brackets indicate it for *q*=6, *q* being the number of levels of the within-subjects variable. The dash (-) indicates theoretical power over .90. The rows for n_j=10 and 16 do not appear because in all the columns of the table the theoretical power was greater than .90; Pw.R= Power Reference of the test.

and negative serial correlation design as the AR matrices, but displaying arbitrary within-subject heterogeneity, so that the variances vary through q without any defined structure, and which were constructed with two deviations from sphericity (ε=.50 and ε=.75). The UN matrices are also arbitrary non-stationary symmetrical, so that both the variances and covariances vary without any defined structure. They lack serial correlation but display absence of sphericity. We constructed matrices with moderate (ε=0.75) and severe (ε=0.56) deviation from sphericity by means of the algorithm developed by Cornell, Young and Bratcher (1991).

While the procedures examined are based on the normality assumption, when we work with real data it is common to find that the skewness (γ_1) and kurtosis (γ_2) indices divert from zero (Micceri, 1989), which could lead us to an incorrect interpretation of the results obtained. Therefore, in order to explore the possible effects of distribution shape on the power of the tests, we carried out a second study in which the assumption of normality underlying the data was violated. Specifically, we chose to investigate the distributions Laplace or double exponential distribution (γ_1=0 and γ_2=3); exponential distribution (γ_1=2 and γ_2=6); and lognormal distribution (γ_1=6.18 and γ_2=110.94). In this second study we manipulated the same variables as in the first study.

In each of the two studies we made comparisons with regard to the accuracy of the estimations and with regard to the empirical power of the three procedures for detecting the within-subject main effect. In order to estimate the degree of bias of the parameters ρ, ε, γ_1 and γ_2 , we compared the mean value of the parameters used, assuming that the estimator is unbiased, so that the mean value will closely approximate the true value of the population parameter. The empirical power rate was calculated by dividing the number of times the null hypothesis is correctly rejected at the specified α level by the number of executions. For the calculation of the theoretical power in those situations where there is absence of sphericity we followed the recommendations of Muller and Barton (1989). Table 1 shows the values of theoretical power of the test.

Generation of data. In order to explore the possible effects of distribution shape on the robustness of the tests, we generated data, both normal and non-normal in form, sampling from the g-and-h distributions introduced by Tukey (1977). Specifically, apart from the standard normal distribution ($g = h = 0$; $\gamma_1 = \gamma_2 = 0$), we chose to investigate distributions where (a) g = 0 and h = .109, a distribution with skew and kurtosis equal to that for a Laplace or double exponential distribution d_1= (γ_1 = 0 and γ_2 = 3); (b) g = .76 and h = –.098, a distribution with skew and kurtosis equal to that for an exponential distribution d_2= (γ_1 = 2 and γ_2 = 6); and (c) g = 1 and h = 0, a distribution with skew and kurtosis equal to that for a lognormal distribution d_3= (γ_1 = 6.18 and γ_2 = 110.94); The g-and-h distributions were obtained by transforming the standard normal variable Z_{ijk} generated using the algorithm proposed by Kinderman and Ramage (1976) by means of the GAUSS program (V. 3.2.32), to $Z^*_{ijk} = g^{-1}[\exp(g\, Z_{ijk})-1]\exp(h\, Z^2_{ijk}/2)$ where g and h are real numbers controlling the skewness and kurtosis, respectively. It should be noted that when g = 0, the g-and-h distribution reduces to $Z^*_{ijk} = Z_{ijk}\exp(h\, Z^2_{ijk}/2)$ which is also known as the h-distribution. Similarly, when h = 0, the g-and-h distribution reduces to $Z^*_{ijk} = [\exp(g\, Z_{ijk})-1]/g$, which is also known as the g-distribution. Finally, The pseudorandom observation vectors $y'_{ij1},\ldots,y'_{ijq}$ with variance-covariance matrix Σ were obtained through triangular decomposition of Σ_j, $Y_{ijk} = T\times(Z^*_{ijk}-\mu_{gh})$ where T is the lower triangular matrix that satisfies the equation Σ_j=TT', and the population mean of the g-and-h distribution is $\mu_{gh}=\{\exp[g^2/(2-2h)-1]\}/[g(1-h)^{1/2}]$ (see Headrick, Kowalchuk, & Sheng, 2008; Wilcox, 1994, for details). Subsequently, using a program written in GAUSS (1992), we carried out as many simulations as experimental conditions described. Each one of these included sampling 5000 independent observations for each of the three procedures.

RESULTS

The tables that follow show the results for a selected subset of studied conditions that adequately show the differences between the different procedures. Table 1, as mentioned earlier, shows the values of theoretical test power for the GG and LEC procedures according to the variables in this study. To highlight the results in the tables (empirical power's smaller than theoretical power's), we took as a reference (Reference Test Power, Pw.R) the mean value of the theoretical test power of the GG procedure GG for n_j=5 and q=4 for all the correlation magnitudes (see Table 1).

Accuracy of the estimations: In each of the situations studied, the values of ε, ρ (positive, negative or zero), γ_1 and γ_2 were systematically adjusted to the parameters of the known populations from those that were simulated.

Estimated Power Rates for the within-subjects main effect:

Normally Distributed Data. Table 2 contains the empirical power for the main effect when data were obtained from a multivariate normal distribution.

1. Stationary autoregressive (AR) and structured non-stationary autoregressive (ARSH-D and ARSH-I) matrices: while it is true that there are differences between the three procedures, these differences are maintained in these three matrix structures. In Table 2 it can be seen that when the serial correlation is positive, the GG, LEC and HCH procedures have an empirical power greater than the theoretical power for all n_j, q and magnitude of ρ when the underlying matrix structure is ARSH-I (results such as those indicated previously are not presented for this reason) and, except when n_j=5 and q=4, also when the underlying deviation matrix is AR and ARSH-D. JN is the procedure with the lowest empirical power.

Table 2
Empirical power for the within-subjects main effect (α=.05). Normal distribution.

q	n_j	ρ	GG q=4	GG q=6	LEC q=4	LEC q=6	HCH q=4	HCH q=6	JN q=4	JN q=6	GG q=4	GG q=6	LEC q=4	LEC q=6	HCH q=4	HCH q=6	JN q=4	JN q=6
			Positive first-order correlation								Negative first-order correlation							
Stationary autoregressive matrix. Matrices AR																		
	5	.20	**.529**	.954	**.574**	.971	**.567**	.972	**.512**	**.949**	**.469**	.960	**.535**	.981	**.523**	.980	**.633**	.993
		.40	**.518**	.910	**.566**	.935	**.559**	.937	**.491**	**.878**	**.347**	.909	**.401**	.942	**.391**	.938	**.671**	.998
		.60	**.510**	.862	**.555**	.891	**.553**	.892	**.435**	**.728**	**.211**	.698	**.244**	.777	**.229**	.757	**.706**	1
		.80	**.505**	.781	**.540**	.813	**.531**	.818	**.410**	**.644**	**.097**	**.242**	**.111**	**.287**	**.103**	**.263**	**.756**	1
	10	.20	.889	1	.899	1	.898	1	.871	.999	.876	1	.888	1	.886	1	.938	1
		.40	.859	.995	.893	.995	.886	.996	.839	.998	.788	.990	.807	.997	.802	1	.954	1
		.60	.837	.992	.869	.974	.869	.994	.783	.985	**.534**	.996	**.559**	.997	**.549**	.996	.972	1
		.80	.834	.980	.844	.982	.844	.993	**.740**	.945	**.205**	**.727**	**.213**	**.747**	**.209**	**.736**	.979	1
	16	.20	.990	1	.991	1	.991	1	.986	1	.987	1	.989	1	.988	1	.995	1
		.40	.980	1	.981	1	.981	1	.967	1	.969	1	.972	1	.971	1	.996	1
		.60	.976	.999	.978	1	.978	.998	.950	.998	.849	1	.858	1	.855	1	.999	1
		.80	.965	.999	.966	.999	.966	.990	.928	.997	**.340**	.989	**.347**	1	**.342**	.990	.999	1
Decreasing structured non-stationary autoregressive matrix. Matrices ARSH-D																		
	5	.20	**.511**	.962	**.560**	.977	**.556**	.977	**.510**	.962	**.423**	.964	**.488**	.983	**.472**	.983	**.594**	.999
		.40	**.499**	.925	**.548**	.949	**.542**	.950	**.460**	.895	**.325**	.930	**.383**	.958	**.368**	.958	**.650**	.999
		.60	**.495**	.875	**.535**	.899	**.534**	.900	**.418**	.798	**.186**	**.735**	**.233**	.803	**.220**	.791	**.693**	1
		.80	**.483**	.792	**.518**	.825	**.510**	.825	**.386**	**.667**	**.104**	**.254**	**.115**	**.296**	**.109**	**.276**	**.727**	1
	10	.20	.878	.999	.893	.999	.892	.999	.859	.999	.842	1	.858	1	.856	1	.918	.999
		.40	.876	.995	.885	.999	.885	.999	.824	.998	.757	1	.776	1	.773	1	.944	1
		.60	.844	.985	.855	.995	.854	.995	**.767**	.987	**.519**	.906	**.542**	.920	**.532**	.916	.961	1
		.80	.811	.978	.822	.986	.823	.987	**.707**	.956	**.170**	.761	**.177**	.781	**.172**	.779	.974	1
	16	.20	.948	1	.986	1	.986	1	.977	1	.983	1	.984	1	.985	1	.992	.999
		.40	.979	1	.980	1	.979	1	.963	1	.958	1	.963	1	.961	1	.996	1
		.60	.971	.999	.973	.999	.973	.999	.943	.999	.823	1	.834	1	.831	1	.997	1
		.80	.959	.999	.962	.999	.962	.999	.913	.996	**.327**	.994	**.332**	.996	**.330**	.996	.999	1
Increasing structured non-stationary autoregressive matrix. Matrices ARSH-I																		
	5	.20	.992	1	.993	1	.995	1	.993	1	.990	1	.996	1	.994	1	.998	1
		.40	.998	1	.993	1	.995	1	.985	1	.972	1	.984	1	.979	1	.999	1
		.60	1	1	1	1	1	1	1	1	**.504**	1	**.510**	1	**.508**	1	.940	1
		.80	1	1	1	1	1	1	1	1	**.423**	1	**.462**	1	**.454**	1	1	1
Arbitrary non-stationary autoregressive matrix. Matrices ARAH																		
q		ε	GG .50	GG .75	LEC .50	LEC .75	HCH .50	HCH .75	JN .50	JN .75	GG .50	GG .75	LEC .50	LEC .75	HCH .50	HCH .75	JN .50	JN .75
	5	.20	**.390**	**.494**	**.416**	**.534**	**.530**	**.573**	**.500**	**.539**	**.355**	**.417**	**.378**	**.465**	**.490**	**.515**	**.583**	**.631**
		.40	**.422**	**.561**	**.444**	**.605**	**.555**	**.631**	**.490**	**.553**	**.336**	**.406**	**.361**	**.458**	**.450**	**.487**	**.670**	**.752**
		.60	**.523**	**.690**	**.543**	**.719**	**.635**	**.731**	**.546**	**.643**	**.296**	**.363**	**.324**	**.397**	**.377**	**.409**	.840	.876
		.80	**.672**	.896	**.688**	.910	.749	.913	**.695**	.836	**.289**	**.292**	**.313**	**.317**	**.309**	**.397**	.976	.980
	10	.20	**.762**	.852	.770	.860	.859	.885	.832	.852	**.734**	.846	.742	.859	.852	.893	.912	.942
4		.40	.808	.902	.814	.910	.887	.992	.836	.876	**.727**	.876	**.738**	.890	.834	.912	.963	.989
		.60	.874	.959	.879	.961	.922	.964	.874	.934	**.740**	.807	.751	.823	.802	.855	.998	.996
		.80	.945	.997	.947	.997	.963	.998	.947	.996	**.699**	**.690**	**.708**	**.711**	**.704**	**.712**	1	1
	16	.20	.954	.975	.955	.976	.986	.983	.979	.960	.955	.986	.958	.988	.985	.992	.993	.998
		.40	.969	.989	.970	.989	.989	.990	.979	.985	.957	.994	.959	.995	.986	.997	.999	1
		.60	.985	.995	.986	.995	.993	.997	.987	.995	.980	.979	.971	.981	.978	.994	1	1
		.80	.996	1	.996	1	.998	1	.996	1	.966	.963	.968	.968	.966	.969	1	1
Non-stationary matrix with arbitrary correlation. Matrices UN (ρ=0)																		
		ε	.56	.75	.56	.75	.56	.75	.56	.75								
4	5		**.236**	**.383**	**.252**	**.414**	**.334**	**.482**	**.324**	**.468**								
	10		**.442**	**.713**	**.452**	**.724**	**.562**	.793	**.544**	.772								
	16		.661	.911	.665	.914	.772	.947	.750	.939								
6	5		.837	.953	.869	.968	.934	.982	.932	.981								
	10		.996	.992	.997	1	.996	1	.999	1								

Note. JN= Jones test (1985); ρ= Correlation in absolute value. For each row, figures in bold indicate empirical power lower than the empirical power of reference. For the rest, see Table 1.

It can also be seen that when the serial correlation is negative, the empirical power of the GG, LEC and HCH procedures is much lower, to a greater degree the greater the serial correlation. In this situation the three previous procedures attain an empirical power greater than the theoretical one when q=8 (if ρ≤.60) for n_j=5. The JN procedure is in this situation the most powerful, the more so the greater the ρ.

2. Arbitrary non-stationary autoregressive (ARAH) and unstructured (UN) matrices: In Table 2 it can be seen that when the underlying matrix in the data is ARAH and the

serial correlation is positive, all the procedures have an empirical power greater than the theoretical one when the within-subject variable has 4 levels only where $n_j \geq 10$. When the serial correlation is negative, the empirical power is greatly inferior for GG, LEC and HCH, but not for the JN procedure, which increases its power as ρ increases.

When the underlying matrix is NE, the empirical power of the four procedures is greater than the theoretical power of reference for q=4, only where ε=.56 and n_j=16 and where ε=.75 and $n_j \geq 10$. The greatest test power is observed in the JN procedure.

Under both matrix structures, ARAH and NE, the empirical power is greater the greater the ε.

Non-normally Distributed Data. Tables 3, 4 and 5 show the empirical power for the main effect with the data obtained from a non-normal distribution.

Table 3

Empirical power for the within-subjects main effect. Covariance structure: AR (α=.05). No normal distribution.

			GG				LEC				HCH				JN			
n_j	*d*	ρ	*q*=4	*q*=6	*q*=8	*q*=12	*q*=4	*q*=6	*q*=8	*q*=12	*q*=4	*q*=6	*q*=8	*q*=12	*q*=4	*q*=6	*q*=8	*q*=12
5	*d*1	.20	**.385**	.815	.988	1	**.436**	.864	.993	1	**.439**	.872	.995	1	**.398**	.819	.987	1
		.80	**.376**	**.607**	.816	.979	**.407**	**.651**	.849	.986	**.405**	**.651**	.853	.987	**.299**	**.479**	.672	.931
	*d*2	.20	**.361**	.775	.977	1	**.404**	.825	.987	1	**.421**	.843	.991	1	**.380**	.791	.977	1
		.80	**.313**	**.580**	.779	.968	**.349**	**.623**	.817	.976	**.353**	**.630**	.825	.980	**.241**	**.445**	**.629**	.901
	*d*3	.20	**.150**	**.348**	**.598**	.903	**.176**	**.409**	**.664**	.924	**.211**	**.482**	**.676**	.972	**.188**	**.419**	**.681**	.952
		.80	**.095**	**.219**	**.361**	**.642**	**.112**	**.258**	**.414**	.689	**.129**	**.287**	**.453**	.730	**.088**	**.174**	**.264**	**.484**
10	*d*1	.20	**.750**	.993	1	1	**.765**	.994	1	1	**.769**	.995	1	1	**.723**	.991	1	1
		.80	**.681**	.918	.988	1	**.692**	.925	.989	1	**.693**	.927	.991	1	**.579**	.822	.951	.999
	*d*2	.20	**.716**	.990	1	1	**.732**	.992	1	1	**.741**	.994	1	1	**.693**	.986	1	1
		.80	**.661**	.907	.980	.999	**.673**	.915	.982	.999	**.679**	.918	.985	.999	**.533**	.804	.936	.997
	*d*3	.20	**.318**	**.660**	.896	.991	**.334**	**.683**	.905	.992	**.369**	**.738**	.945	.998	**.332**	**.677**	.911	.996
		.80	**.240**	**.485**	**.693**	.907	**.253**	**.502**	.710	.913	**.277**	**.537**	**.739**	.929	**.189**	**.346**	**.516**	.786
16	*d*3	.20	**.485**	.856	.976	.998	**.497**	.863	.977	.998	**.528**	.893	.989	.999	**.480**	.851	.981	.999
		.80	**.417**	**.700**	.870	.974	**.424**	.709	**.875**	.975	**.447**	**.732**	.888	.980	**.320**	**.542**	**.743**	.930
5	*d*1	-20	**.319**	.819	.993	1	**.369**	.874	.997	1	**.368**	.882	.998	1	**.476**	.948	1	1
		-40	**.235**	**.711**	.985	1	**.282**	.787	.993	1	**.276**	.792	.994	1	**.512**	.975	1	1
		-60	**.151**	**.472**	.895	.999	**.180**	**.552**	.935	.999	**.173**	**.547**	.936	1	**.555**	.989	1	1
		-80	**.082**	**.159**	**.397**	.956	**.090**	**.192**	**.468**	.973	**.085**	**.177**	**.444**	.974	**.589**	.996	1	1
	*d*2	-20	**.303**	**.763**	.982	1	**.350**	.827	.991	1	**.365**	.845	.995	1	**.459**	.930	.999	1
		-40	**.241**	**.660**	.964	1	**.281**	.733	.981	1	**.283**	.751	.985	1	**.496**	.960	.999	1
		-60	**.168**	**.434**	.826	.999	**.195**	**.502**	.878	1	**.193**	**.505**	.886	1	**.544**	.979	1	1
		-80	**.126**	**.189**	**.383**	.988	**.139**	**.217**	**.442**	.939	**.137**	**.212**	**.430**	.940	**.601**	.992	1	1
	*d*3	-20	**.130**	**.309**	**.573**	.999	**.159**	**.373**	**.646**	.930	**.191**	**.454**	**.748**	.979	**.242**	**.582**	.858	.992
		-40	**.114**	**.241**	**.481**	.877	**.137**	**.303**	**.566**	.911	**.160**	**.364**	**.653**	.964	**.271**	**.660**	.906	.995
		-60	**.105**	**.162**	**.315**	**.756**	**.119**	**.201**	**.387**	.808	**.132**	**.227**	**.437**	.876	**.310**	**.708**	.840	.998
		-80	**.113**	**.109**	**.144**	**.372**	**.120**	**.125**	**.176**	**.443**	**.130**	**.134**	**.190**	**.477**	**.355**	.775	.959	.991
10	*d*1	-20	**.693**	.997	1	1	**.711**	.997	1	1	**.716**	.998	1	1	.808	.999	1	1
		-40	**.571**	.991	1	1	**.595**	.994	1	1	**.595**	.993	1	1	.842	.999	1	1
		-60	**.355**	.942	.999	1	**.376**	.952	.999	1	**.368**	.952	.999	1	.880	1	1	1
		-80	**.132**	**.449**	.930	1	**.138**	**472**	.941	1	**.134**	**.464**	.941	1	.899	1	1	1
	*d*2	-20	**.643**	.991	1	1	**.664**	.993	1	1	**.672**	.995	1	1	.768	.999	1	1
		-40	**.532**	.979	1	1	**.555**	.983	1	1	**.559**	.986	1	1	.819	.999	1	1
		-60	**.347**	.889	.999	1	**.364**	.905	.999	1	**.363**	.907	.999	1	.847	.999	1	1
		-80	**.165**	**.424**	.866	1	**.170**	**.445**	.880	1	**.170**	**.438**	.883	1	.882	1	1	1
	*d*3	-20	**.277**	**.639**	.890	.991	**.292**	**.666**	.903	.992	**.325**	**.728**	.946	.998	**.402**	.828	.978	.999
		-40	**.230**	**.548**	.847	.988	**.245**	**.578**	.865	.989	**.269**	**.634**	.913	.997	**.449**	.888	.989	.999
		-60	**.165**	**.367**	**.693**	.971	**.174**	**.391**	**.718**	.974	**.187**	**.430**	**.766**	.988	**.492**	.920	.993	.999
		-80	**.128**	**.173**	**.315**	.787	**.134**	**.183**	**.337**	.808	**.138**	**.192**	**.359**	.834	**.541**	.945	.997	1
16	*d*1	-20	.910	1	1	1	.916	1	1	1	.917	1	1	1	.955	1	1	1
		-40	.847	1	1	1	.857	1	1	1	.958	1	1	1	.971	1	1	1
		-60	**.643**	.999	1	1	**.653**	.999	1	1	.653	.999	1	1	.981	1	1	1
		-80	**218**	.837	.999	1	**.223**	.845	.999	1	**.221**	.843	.999	1	.988	1	1	1
	*d*2	-20	.875	.999	1	1	.882	.999	1	1	.885	.999	1	1	.938	1	1	1
		-40	.796	.999	1	1	.807	.999	1	1	.811	1	1	1	.956	1	1	1
		-60	**.580**	.995	1	1	**.591**	.995	1	1	**.591**	.996	1	1	.971	1	1	1
		-80	**.246**	**.733**	.995	1	**.249**	**.745**	.996	1	**.249**	**.743**	.996	1	.981	1	1	1
	*d*3	-20	**.421**	.843	.979	.998	**.433**	.852	.980	.999	**.466**	.889	.992	.999	**.560**	.944	.997	1
		-40	**.339**	.771	.964	.998	**.351**	.784	.967	.998	**.374**	.823	.983	.999	**.603**	.969	.998	1
		-60	**.234**	**.596**	.903	.995	**.241**	**.611**	.909	.995	**.253**	**.640**	.934	.998	**.651**	.981	.999	1
		-80	**.142**	**.248**	.539	.952	**.144**	**.256**	**.554**	.955	**.150**	**.264**	**.572**	.966	**.701**	.986	.999	1

Note. *d*: *d*1= (γ_1=0, γ_2=3); *d*2= (γ_1=2, γ_2=6); *d*3= (γ_1=6.18, γ_2=110.94). For the rest, see Tables 1 and 2.

Table 4

Empirical power for the within-subjects main effect. Covariance structure: ARSH-I. (α=.05). No normal distribution.

			GG				LEC				HCH				JN			
n_j		ρ	*q*=4	*q*=6	*q*=8	*q*=12	*q*=4	*q*=6	*q*=8	*q*=12	*q*=4	*q*=6	*q*=8	*q*=12	*q*=4	*q*=6	*q*=8	*q*=12
5	*d*1	.20	.934	.997	.999	1	.947	.948	.999	1	.955	.999	1	1	.944	.998	.999	1
		.80	.854	.845	.980	.997	.867	.953	.984	.998	.882	.962	.988	.999	.820	.911	.957	.991
	*d*2	.20	.958	.999	1	1	.970	1	1	1	.977	1	1	1	.962	1	1	1
		.80	.899	.980	.993	.999	.913	.984	.996	.999	.935	.991	.998	.999	.862	.956	.978	.998
	*d*3	.20	**.527**	.795	.915	.979	**.572**	.828	.933	.983	**.648**	.909	.980	.998	**.590**	.861	.962	.997
		.80	**.357**	**.531**	**.661**	.833	**.394**	**.575**	**.710**	.867	**.458**	**.657**	.782	.918	**.321**	**.428**	**.511**	**.661**
10	*d*1	.20	.999	1	1	1	.999	1	1	1	.999	1	1	1	.999	1	1	1
		.80	.991	.999	1	1	.991	.999	1	1	.993	.999	1	1	.986	.998	1	1
	*d*2	.20	.999	1	1	1	.999	1	1	1	.999	1	1	1	.999	1	1	1
		.80	.999	1	1	1	.999	1	1	1	1	.998	.998	1	.998	1	1	1
	*d*3	.20	.880	.976	.993	.998	.889	.978	.994	.998	.918	.992	.999	1	.889	.986	.998	.999
		.80	.821	.931	.976	.994	.831	.937	.978	.994	.874	.966	.989	.998	.747	.854	.921	.976
5	*d*1	-20	.912	.998	.999	1	.930	.998	1	1	.943	.999	1	1	.970	1	1	1
		-40	.852	.995	.999	1	.878	.997	.999	1	.894	.998	1	1	.981	1	1	1
		-60	**.656**	.962	.998	1	**.691**	.972	.998	1	**.712**	.982	.999	1	.982	1	1	1
		-80	**.288**	**.600**	.865	.995	**.312**	**.640**	.893	.997	**.321**	**.666**	.910	.998	.989	1	1	1
	*d*2	-20	.917	.999	1	1	.938	.999	1	1	.951	.999	1	1	.982	1	1	1
		-40	.820	.996	1	1	.849	.998	1	1	.876	.999	1	1	.986	1	1	1
		-60	**.607**	.942	.997	1	**.643**	.957	.998	1	**.666**	.972	.999	1	.990	1	1	1
		-80	**.296**	**.549**	.808	.990	**.314**	**.587**	.841	.993	**.330**	**.612**	.866	.997	.992	1	1	1
	*d*3	-20	**.460**	**.752**	.896	.971	**.501**	.796	.918	.976	**.570**	.887	.976	.998	.680	.646	.991	.999
		-40	**.399**	**.652**	.838	.957	**.407**	**.704**	.872	.965	**.469**	.798	.945	.994	.726	.965	.995	.999
		-60	**.269**	**.467**	**.663**	.897	**.295**	**.517**	**.712**	.918	**.332**	**.593**	.805	.972	.769	.973	.996	.999
		-80	**.177**	**.220**	**.307**	**.551**	**.188**	**.246**	**.346**	**.606**	**.203**	**.277**	**.396**	**.680**	.786	.979	.996	.999
10	*d*1	-20	.999	1	1	1	.999	1	1	1	.999	1	1	1	.999	1	1	1
		-40	.998	1	1	1	.999	1	1	1	.999	1	1	1	.999	1	1	1
		-60	.982	1	1	1	.984	1	1	1	.988	1	1	1	.999	1	1	1
		-80	**.715**	.990	.999	1	**.726**	.991	.999	1	**.747**	.993	1	1	.999	1	1	1
	*d*2	-20	.999	1	1	1	.999	1	1	1	1	1	1	1	1	1	1	1
		-40	.998	1	1	1	.998	1	1	1	.998	1	1	1	1	1	1	1
		-60	.968	1	1	1	.971	1	1	1	.976	1	1	1	1	1	1	
		-80	**.640**	.971	.999	1	**.651**	.974	.999	1	**.671**	.979	.999	1	1	1	1	1
	*d*3	-20	.813	.969	.991	.997	.826	.971	.991	.997	.871	.991	.998	.999	.928	.995	.999	1
		-40	**.715**	.943	.984	.997	**.728**	.950	.985	.997	.775	.976	.996	.999	.937	.996	.999	1
		-60	**.519**	.843	.951	.991	**.533**	.853	.956	.991	**.576**	.901	.980	.998	.950	.998	.999	1
		-80	**.268**	**.443**	**.685**	.931	**.273**	**.459**	**.701**	.938	**.295**	**.505**	**.756**	.965	.949	.998	.999	1
16	*d*3	-20	.948	.944	.998	999	.951	.995	.998	.999	.967	.999	.999	.999	.985	1	.999	.999
		-40	.902	.990	.997	.999	.906	.991	.997	.999	.929	.996	.999	.999	.989	.999	.999	.999
		-60	**.757**	.968	.992	.999	**.763**	.969	.992	.999	.796	.982	.997	.999	.989	.999	1	1
		-80	**.411**	**.726**	.914	.999	**.416**	**.734**	.917	.999	**.442**	.770	.940	.999	.988	.999	1	1

Note. See Tables 1, 2 and 3

1. Stationary autoregressive (AR) and decreasing structured non-stationary autoregressive (ARSH-D) matrices: Behaviour of the four procedures is highly similar when the data are underlain by two matrix structures, the only difference being that the empirical power is slightly less in ARSH-D.

In Table 3 it can be seen that when the serial correlation is positive, none of the four procedures has an empirical power greater than the theoretical power of reference when q=4, n_j≤10 in any of the non-normal distributions studied. For the test power to be greater than that of reference it is necessary, in addition to the data distribution not diverging too much from normality, that when n_j=5 the number of levels of the within-subject variable to be 8 or more, and if n_j=10, q must be 6 or more.

When the serial correlation is negative the power is lower for all procedures, except JN, which when n_j=10 attains a test power greater than that of reference in all q under the distributions d_1 and d_2.

2. Increasing structured non-stationary autoregressive (ARSH-I) matrix: In Table 4 it can be seen that when the serial correlation is positive the four procedures exhibit very high empirical power for all ρ, q and n_j when the distribution underlying the data is d_1 and d_2. When the serial correlation is negative, the JN procedure shows excellent behaviour, and the GG, LEC and HCH procedures, except when the correlation magnitude is high and n_j=5, also show power greater than that of reference.

3. Arbitrary non-stationary autoregressive (ARAH) matrix: Table 5 shows that when the serial correlation is posi-

Table 5

Empirical power for the within-subjects main effect. Covariance structure: ARAH and NE (α=.05). No normal distribution.

Arbitrary non-stationary autoregressive matrix. Matrices ARAH																		
			GG				LEC				HCH				JN			
		q	4		8		4		8		4		8		4		8	
		ε	.50	.75	.50	.75	.50	.75	.50	.75	.50	.75	.50	.75	.50	.75	.50	.75
n_j	*d*	ρ																
5	*d*1	.20	**.316**	**.364**			**.336**	**.398**			.428	.437			**.406**	**.398**		
		.40	**.338**	**.423**	.961	.999	**.356**	**.463**	.972	.994	**.448**	**.489**	.987	.996	**.390**	**.413**	.951	.989
		.60	**.404**	**.542**	.978		**.425**	**.577**	.984		**.503**	**.589**	.990		**.425**	**.499**	.958	
		.80	**.519**	**.760**			**.538**	.785			**.601**	.786			**.542**	**.677**		
	*d*2	.20	**.202**	**.283**			**.228**	**.321**			**.346**	**.385**			**.315**	**.346**		
		.40	**.233**	**.354**	.986	.999	**.259**	**.397**	.991	.999	**.366**	**.442**	.998	1	**.307**	**.353**	.981	.999
		.60	**.313**	**.493**	.995		**.337**	**.530**	.996		**.439**	**.552**	.999		**.357**	**.455**	.984	
		.80	**.446**	**.727**			.472	.752			**.562**	**.761**			**.495**	**.657**		
	*d*3	.20	**.059**	**.074**			**.071**	**.093**			**.121**	**.136**			**.120**	**.125**		
		.40	**.065**	**.092**	**.538**	**.615**	**.077**	**.116**	**.614**	**.680**	**.126**	**.150**	.789	.790	**.124**	**.123**	**.581**	**.642**
		.60	**.074**	**.187**	**.603**		.088	.217	.668		**.134**	**.250**	.803		**.126**	**.199**	**.539**	
		.80	**.114**	**.410**			**.128**	**.435**			**.186**	**.458**			**.162**	**.358**		
10	*d*1	.20	**.591**	**.692**			**.598**	**.706**			**.701**	**.748**			**.667**	**.705**		
		.40	**.642**	**.756**	.999	.999	**.650**	**.768**	.999	1	.748	.794	.999	1	**.679**	.725	1	1
		.60	**.728**	.868	1		**.735**	.874	1		.800	.884	1		**.730**	.820	1	
		.80	.835	.973			.838	.976			.877	.976			.835	.948		
	*d*2	.20	**.581**	**.677**			**.595**	**.695**			**.736**	**.739**			**.681**	**.687**		
		.40	**.633**	**.756**	1	1	**.645**	**.769**	1	1	.768	.799	1	1	**.681**	**.714**	1	1
		.60	**.730**	.852	1		**.739**	.862	1		.827	.874	1		**.734**	**.743**	1	
		.80	.859	.947			.863	.950			.905	.954			.866	.921		
	*d*3	.20	**.150**	**.218**			**.159**	**.235**			**.260**	**.306**			**.225**	**.263**		
		.40	**.165**	**.268**	.938	.932	**.174**	**.281**	.945	.940	**.275**	**.345**	.987	.972	**.225**	**.266**	.930	.941
		.60	**.216**	**.421**	.996		**.225**	**.437**	.970		**.324**	**.480**	.991		**.257**	**.386**	.933	
		.80	**.332**	**.661**			**.341**	**.672**			.435	.691			**.369**	**.590**		
16	*d*3	.20	**.298**	**.408**			**.306**	**.418**			**.455**	**.490**			**.389**	**.436**		
		.40	**.340**	**.481**	.993	.989	**.348**	**.494**	.993	.989	**.489**	**.553**	.993	1	**.397**	**.450**	.994	1
		.60	**.443**	**.634**	.995		**.450**	**.641**	.995		**.568**	**.674**	.999		**.451**	**.568**	.994	
		.80	**.585**	.800			**.590**	.804			**.683**	.819			**.608**	**.742**		
5	*d*1	-20	**.263**	**.297**			**.285**	**.338**			.376	.380			**.457**	**.479**		
		-80	**.204**	**.209**			**.224**	**.228**			**.218**	**.223**			.898	.903		
	*d*2	-20	**.162**	**.216**			**.181**	**.253**			**.298**	**.322**			**.397**	**.428**		
		-80	**.256**	**.252**			**.272**	**.268**			.271	.269			.887	.881		
	*d*3	-20	**.053**	**.060**			**.061**	**.074**			**.114**	**.117**			**.141**	**.156**		
		-80	.183	.189			**.195**	**.201**			**.206**	**.212**			**.572**	**.577**		
10	*d*1	-20	**.545**	**.648**			.556	.667			**.681**	**.719**			.765	.806		
		-80	**.474**	**.479**			**.485**	**.490**			**.489**	**.493**			.997	.996		
	*d*2	-20	**.496**	**.614**			**.513**	**.638**			**.688**	**.705**			.806	.824		
		-80	**.458**	**.454**			.465	.466			**.468**	**.465**			.996	.997		
	*d*3	-20	**.117**	**.163**			.122	.178			**.227**	**.249**			.296	.330		
		-80	**.253**	**.252**			**.258**	**.256**			**.268**	**.268**			.796	.797		
16	*d*3	-20	**.232**	.323			**.238**	**.334**			**.395**	**.424**			**.516**	**.541**		
		-80	**.239**	**.335**			**.334**	**.340**			.341	.349			.924	.924		

Non-stationary matrix with arbitrary correlation. Matrices UN (ρ=0)																	
		ε =.56(n_j=5)				ε =.75(n_j=5)				ε =.56(n_j=10)				ε =.75(n_j=10)			
d	*q*	GG	LEC	HCH	JN	GG	LEC	HCH	JN	GG	LEC	HCH	JN	GG	LEC	HCH	JN
*d*1	4	**.174**	**.188**	**.259**	**.251**	**.274**	**.298**	**.363**	**.351**	**.329**	**.335**	**.442**	**.420**	**.540**	**.552**	**.628**	**.607**
	6	.664	.714	.804	.806	.820	.856	.897	.905	.961	.965	.987	.986	.994	.995	.997	.997
*d*2	4	**.137**	**.148**	**.230**	**.216**	**.232**	**.263**	**.338**	**.326**	**.295**	**.303**	**.413**	**.391**	**.511**	**.526**	**.616**	**.594**
	6	.643	.693	.795	.792	.821	.862	.906	.906	.953	.960	.981	.978	.994	.995	.998	.998
*d*3	4	**.071**	**.078**	**.136**	**.128**	**.098**	**.113**	**.171**	**.161**	**.122**	**.125**	**.204**	**.185**	**.205**	**.214**	**.298**	**.277**
	6	**.266**	**.314**	**.436**	**.447**	**.359**	**.414**	**.530**	**.545**	**.540**	**.564**	.687	.683	**.696**	**.716**	.797	.798
	8	**.495**	**.560**	.709	.702	**.617**	**.682**	.804	.802	.837	.853	.926	.917	.920	.929	.966	.963

Note. See Tables 1, 2 and 3

tive and *q*=4, all four procedures attain a test power greater than that of reference when ρ=.80, n_j=10, and the distribution underlying the data is d_1 and d_2. When the correlation is negative, only the JN procedure attains high test powers. If n_j=16, the test power is always high, whether the serial correlation is positive or negative.

When q=8 all the procedures have high empirical power except when the distribution underlying the data is d_3. The empirical power is slightly higher when ε=.75.

4. Unstructured (UN) matrix: Table 5 shows that when the distribution underlying the data is d_1 and d_2 all the procedures have an empirical power greater than that of reference when q≥6 for all n_j. The test power is much lower in d_3. The empirical power is slightly higher when ε=.75.

DISCUSSION AND RECOMMENDATIONS

The objective of the present research was to compare the behaviour of the univariate procedures GG, LEC, HCH and JN with regard to the power rates in the additive model of a split-plot factorial design ($p \times q$). Two aspects help to make it original. On the one hand, the comparison of procedures that differ in the way they deal with absence of sphericity, either assuming arbitrary correlation or assuming serial autocorrelation. On the other hand, the underlying conditions in the data added to the absence of sphericity under which the behaviour was compared, including, in addition to those commonly studied, the arbitrary non-stationary autoregressive condition that permits us to observe what affects the empirical power most – the absence of sphericity or the direction and amount of autocorrelation. The results highlight the following:

Differences found according to the distribution underlying the data:

1. When the data are underlain by a normal distribution and serial correlation is positive, the LEC and HCH procedures are those exhibiting the greatest empirical power, and the JN procedure is that with the least empirical power when the underlying deviation matrix is AR and ARSH. When the underlying deviation matrix is ARAH and NE the most powerful procedures are HCH and JN if the serial correlation is negative, and the JN procedure is indisputably that with the greatest power whatever the underlying deviation matrix is.

2. When the data are underlain by a non-normal distribution, the HCH procedure is that with the greatest empirical power when the serial correlation is positive, and the JN procedure when the serial correlation is negative whatever the underlying deviation matrix is. The GG procedure GG is that which has demonstrated the lowest empirical power in all the deviation matrix structures studied here.

Coincident behaviour patterns when the data are distributed normally and non-normally:

1. In all three procedures the empirical power increases as q, n_j and ε increase, more with q than with n_j; and with n_j more than with ε.

2. An increase in the magnitude of the correlation always affects the estimation of the test power in all procedures; however, the following was observed:

a) When the matrix underlying the data is AR, ARSH-I and ARSH-D, when the serial correlation is positive, an increase in magnitude leads to a reduction in power in all the procedures; however, no significant change is found between the estimation for high and low levels of serial correlation. The most appreciable changes are found when q=6 and 8 for n_j=5. On the other hand, when the serial correlation is negative, an increase in magnitude leads to a significant reduction in power in the GG, LEC and HCH procedures, and an increase in power (also significant) in JN. Both changes are larger when q=8 in all sample sizes.

Previously, in point 1, it was stressed that the empirical power, in relation to ε=.50, is slightly higher when ε=.75. This point now needs some clarification. A look at Table A –which shows the sphericity coefficients in these types of matrix structure for all correlation magnitudes studied– reveals that when the matrices have positive serial correlation, the values of ε decrease very gradually as the magnitude of ρ increases; however, when the serial correlation is negative, the values of ε decrease markedly as the magnitude of ρ increases. This is the reason for the behaviour of the test power. Thus, it would appear that estimation of the empirical test power is determined more by the magnitude of the deviation from sphericity than by the magnitude of the serial correlation when the underlying deviation matrix is AR and ARSH.

b) When the matrix underlying the data is ARAH, if the serial correlation is positive, an increase in magnitude leads to a significant increase in test power in all the procedures. When the correlation is negative, the test power decreases almost imperceptibly as the magnitude of ρ increases, except in the case of the JN procedure, whose estimation improves as ρ increases. Thus, it would appear that estimation of the empirical test power is determined more by the magnitude of the correlation than by the deviation from sphericity when the underlying deviation matrix is positive ARAH.

3. According to the structure of deviation underlying the data, the magnitude of the empirical power for all the procedures is as follows:

$(1-\hat{\beta})$ARSH-I > $(1-\hat{\beta})$AR > $(1-\hat{\beta})$ARSH-D > ARAH > NE.

That is, if the matrix is severely increasing structured non-stationary autoregressive, all three procedures have very high empirical power, and much higher then the theoretical power for all n_j, q, and magnitude and direction of the autocorrelation.

4. The greater the deviation from normality, the lower the empirical test power, as follows:

When the deviation matrix underlying the data is AR, ARSH-D, ARAH and NE:

$(1-\hat{\beta})$ Normal distribution > $(1-\hat{\beta})\ d_1$ > $(1-\hat{\beta})\ d_2$ > $(1-\hat{\beta})\ d_3$.

When the deviation matrix underlying the data is ARSH-I:

$(1-\hat{\beta})$ Normal distribution > $(1-\hat{\beta})\ d_2$ > $(1-\hat{\beta})\ d_1$ > $(1-\hat{\beta})\ d_3$.

The results we have obtained with the GG and LEC procedures replicate those found in the numerous previous studies in similar conditions, that is, AR, ARSH and UN matrices (the reader may wish to consult the literature covered in the reviews referred to here). Nevertheless, the fact of having compared these tests with the HCH and JN procedures, and of having the ARAH matrix structure added to the study, broadened the scope of behaviour of these procedures that correct deviation from sphericity based on different criteria.

Despite the fact that the generality of our results is limited by the range of conditions and parameter sets employed in the simulations –other conditions or other parameter sets could give different results–, in our opinion, when all subjects have complete response vectors a general recommendation can be made. Applied researchers who in gathering their data employ a repeated-measures design with a within-subject and a between-subjects variable, and have manipulated the levels of the variables in order to select those that are optimum for testing the within-subject treatment effects (they have to obtain at least 10 subjects for each of the groups if they have a number of repeated-measures of 4 or less; if the within-subject variable has more than 4 levels, five subjects in each group would be sufficient), must first of all study whether or not the assumptions have been met. If after doing so they reach the conclusion that the assumption of sphericity is not met, and if they decide to use a univariate statistic, here are suggestions for testing the within-subject treatment effects:

- If there is positive serial correlation and the data are normally distributed, LEC or HCH Statistics would be the best choice; if there is positive serial correlation and the data are non-normally distributed, HCH Statistic would be the best choice.
- If there is negative serial correlation and either normal or non-normal distribution of data, JN Statistic would be the best choice.

As a final note, we should point out the need to continue research on the behaviour of these procedures with respect to Type I error and to test power in the situations studied here when the sizes of the subsamples of units of study produce an unbalanced design, and when the covariance matrices of the groups are heterogeneous.

REFERENCES

Blanca Mena, M. (2004). Alternativas de análisis estadístico en los diseńos de medidas repetidas [Approaches to the statistical analysis of repeated measures designs]. *Psicothema, 16(3),*509-518.

Collier, R. O., Baker, F. B., Mandeville, G. K., & Hayes, T. F. (1967). Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures designs. *Psychometrika, 32,* 339-353.

Cornell, J.E., Young, D.M., & Bratcher, T.L. (1991). An algorithm for generating covariance matrices with specified departures from sphericity. *Journal of Statistical Computation and Simulation, 34*, 240-243.

Edwards, L.K. (1991). Fitting a serial correlation pattern to repeated observations. *Journal of Educational Statistics, 16* (1), 53-76.

Fernández, P. (1995). *Comparación mediante simulación de la potencia y robustez del enfoque multivariado de medidas repetidas frente al correspondiente univariado con la estructura del error modelada a través de procesos autorregresivos* [*Comparison by simulation of the power and robustness of multivariate repeated measures approach compared to the corresponding univariate with the error structure modeled by autoregressive processes*]. Unpublished doctoral disserataion. Universidad de Oviedo.

Fernández, P., Livacic-Rojas, P., & Vallejo, G. (2007). Cómo elegir la mejor prueba estadística para analizar un diseńo de medidas repetidas [Choosing the best statistical test to analyze a repeated measures design]. *International Journal of Clinical and Health Psychology, 7*(1), 153-175.

Fernández, P., Vallejo, J., Livacic-Rojas, Herrero, J., & Cuesta, M. (2008). Comparative robustness of six tests in repeated measures designs with specified departures from sphericity. *Quality & Quantity,* DOI: 10.1007/S11135-008-9198-3.

Fernández, P., Vallejo, G., & Livacic-Rojas, P. (2008). Comparison of the robustness of the SPSS MIXED procedure with regard to another three univariate statistics in repeated measures designs with specified departures from sphericity. *Paper presented at the III European Congress of Methodology*, Oviedo (Spain). July 8-12, 2008.

GAUSS. (1992). *The Gauss System*. (Vers. 3.2.32). Washington: Aptech Systems, Inc.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112.

Headrick, T. C., Kowalchuk, R. K., & Sheng, Y. (2008). Parametric probability densities and distribution functions for Tukey g-and-h transformations and their use for fitting data. *Applied Mathematical Sciences, 2*, 449-462.

Hearne, E.M., Clark, G.M., & Hatch, J.P. (1983). A test for serial correlation in univariate repeated-measures analysis. *Biometrics, 39*, 237-243.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurements design have exact F-distributions. *Journal of the American Statistical Association, 65*, 1582-1585.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data randomized block and split-plot designs. *Journal of Educational Statistics, 1,* 69-82.

Jones, R.H. (1985). Repeated measures, interventions, and time series analysis. *Psychoneuroendocrinology, 10(1)*, 5-14.

Keselman, H. J., Algina, J., & Kowalchuck, R. K. (2001). The analysis of the repeated measures design: A review. *British Journal of Mathematical and Statistical Psychology, 54,* 1-20.

Keselman, H.J., Algina, J., Kowalchuck, R.K., & Wolfinger, R.D. (1999a). The analysis of repeated measurements: A comparison of mixed-model Satterthwaite F tests and a nonpooled adjusted degrees of freedom multivariate test. *Communications in Statistics-Theory and Methods, 28*, 2967-2999.

Keselman, H.J., Algina, J., Kowalchuck, R.K., & Wolfinger, R.D. (1999 b). A comparison of recent approaches to the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology, 52*, 63-78.

Keselman, J., Lix, L., & Keselman H. J. (1996). The analysis of repeated measurements designs: A quantitative research synthesis. *British Journal of Mathematical and Statistical Psychology, 49,* 275-298.

Kinderman, A.J., & Ramage, J.G. (1976). Computer generation of normal random numbers. *Journal of American Statistical Association 77*, 893-896.

Lecoutre, B. (1991). A correction for the $\tilde{\varepsilon}$ approximate test in repeated measures designs with two or more independent groups. *Journal of Educational Statistics, 16*, 371-372.

Livacic, P. (2005). *Una evaluación empírica de procedimientos alternativos de análisis e diseńos de medidas repetidas* [An empirical evaluation of alternative procedures for analysis of repeated measures designs]. Unpublished doctoral dissertation. Universidad de Oviedo.

Méndez, X., Orgiles, M., & Espada, J.P. (2004). Escenificaciones emotivas para la fobia a la oscuridad: un ensayo controlado [Emotional performances for darkness phobia: a controlled trial]. *International Journal of Clinical and Health Psychology 4*(3), 505-520.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 92*, 778-785.

Muller, K.E., & Barton, C.N. (1989). Approximate power for repeated-measures ANOVA lacking sphericity. *Journal of the American Statistical Association, 84* (406), 549-555.

Palmero, F., Breva, A., Diago, J.L., Díez. J.L., & García, I. (2002). Funcionamiento psicofisiológico y susceptibilidad a la sintomatología premestrual en mujeres Tipo A y Tipo B [Psychophysiological functioning and susceptibility to premestrual symptoms in women Type A and Type B]. *International Journal of Clinical and Health Psychology 2*(1), 111-136.

Quintana, S., & Maxwell, S. E. (1994). A Monte Carlo comparison of seven ε-adjustment procedures in repeated measures designs with small sample sizes. *Journal of Educational Statistics, 19*, 57-71.

Tukey, J.W. (1977). Modern techniques in data analysis. NSF-sponsored regional research conference at Southern Massachusetts University (North Dartmouth, MA).

Vallejo, G. (1991). *Análisis univariado y multivariado de los diseńos de medidas repetidas de una sola muestra y de muestras divididas* [Univariate and multivariate repeated measures designs in a single sample and split simples]. Barcelona, PPU.

Vallejo, G., Arnau, J., Bono, R., Cuesta, M., Fernández, P., & Herrero, J. (2002). Análisis de diseńos de series temporales cortas [Analysis of short time series designs]. *Metodología de las Ciencias del Comportamiento, 4*(2), 301-323.

Vallejo, G., Fernández, P., & Ato, M. (2003). Tasas de potencia de dos enfoques robustos para analizar datos longitudinales [Power rates of two robust approaches for to analyze longitudinal data]. *Psicológica, 24,* 109-122.

Wilcox, R. R. (1994) A one-way random effects model for trimmed means. *Psychometrika, 59*, 289-306.