Notes and Comment

Further evidence for Coren and Hakstian's "Methodological implications of interaural correlation: Count heads not ears" and an alternative correction formula

BRUNO D. ZUMBO

University of Ottawa, Ottawa, Ontario, Canada

and

DONALD W. ZIMMERMAN

Carleton University, Ottawa, Ontario, Canada

Coren and Hakstian (1990) identified a serious methodological problem that arises in auditory research because of interaural correlation. When measures from both ears of the subjects are pooled together in an experimental design that assumes independence of measures, there can be spuriously high apparent statistical significance. The present paper provides further evidence in support of Coren and Hakstian's argument and also derives a formula that effectively corrects inflated test statistics resulting from interaural correlation. This formula is a special case of a more general one that applies in many other experimental contexts in which nonindependence of measures is a problem. We found that statistical tests based on our formula have somewhat greater power to detect differences than the kind of correction method advocated by Coren and Hakstian.

Coren and Hakstian (1990) recently identified an important statistical problem that arises in research that assesses the efficiency of limited-frequency testing and screening for auditory sensitivity (see, e.g., Wiley, Oviatt, & Block, 1987). They noted that the practice of treating measures of individual ears of one subject as independent, when in fact they are not, can lead to serious inflation of the Type I error rate during statistical inference. In other words, assuming independence when in fact the observations are not independent increases the probability of Type I errors.

Coren and Hakstian (1990) established empirically that there is a substantial nonzero correlation between auditory sensitivities of the two ears, sometimes as high as .80, and demonstrated that the correlation is not an artifact of bilateral hearing loss occurring with age. The same problem arises in visual research that treats the measures obtained on a subject's two eyes as independent

when they are not and also in laterality research when the interlateral correlation is assumed to be zero.

The purpose of this paper is to present further evidence that Coren and Hakstian's conclusions are indeed correct and relevant to much psychonomic research. In particular, we shall describe some recently developed theoretical derivations that deal with the lack of independence and then further demonstrate the inflation of the probability of Type I error with lack of independence. The focus of this paper will be on a simple case of the two-independent-samples design, sometimes referred to as a completely randomized design with one treatment factor and two groups, although similar reasoning applies to more complex designs.

We will first review some of the findings of Zimmerman, Williams, and Zumbo (1990) and then present some formulas for correcting t values when observations are not independent. The corrected t formulas will then be evaluated by monte carlo simulation intended to model the kind of design described by Coren and Hakstian (1990). Finally, a detailed example of application of the correction formulas will be given.

THE PROBLEM OF NONINDEPENDENCE

For the completely randomized design, most traditional statistical tests, both parametric and nonparametric, assume independence within and between groups.¹ Recently, Zimmerman et al. (1990) studied the effect of violations of independence on the independent samples t test, and the results are consistent with the argument presented by Coren and Hakstian (1990). That is, they reveal that Type I error probabilities can be substantially inflated when observations are not independent. Furthermore, they show that when a modified formula is used to calculate the t statistic, the probabilities are no longer inflated. However, the Zimmerman et al. derivations are based on the general case of exchangeable random variables (i.e., dependent, identically distributed random variables). Exchangeability is a mathematical generalization of the concept of independent, identically distributed random variables. In this case, there is even stronger statistical dependence than in the auditory research design in which measures based on two ears are dependent.

Table 1 demonstrates the notions of independence, exchangeability, and the special dependence in the interaural correlation case. The matrices are symmetrical variance-covariance matrices for two subjects,² where X_1 and X_2 are the random variables for the two ears for Subject 1 and X_3 and X_4 are the random variables for Subject 2. The terms on the diagonal are the variances, and the off-

This research was made possible by a Social Sciences and Humanities Research Council of Canada Fellowship to the first author and by a Carleton University Research Grant to the second author. We would like to express our gratitude to Joseph V. Baranski for comments on an earlier draft of this paper. Send correspondence to Bruno D. Zumbo, Measurement and Evaluation, Faculty of Education, University of Ottawa, 145 Jean-Jacques Lussier, Ottawa, Ont. K1N 6N5, Canada.

Table 1
Variance-Covariance Matrices for Independence, Exchangeability,
and the Interaural Correlation Design

Matrix A				
	X_1	X_2	X_3	X_4
X_1	σ_1^2			
X_2	0	σ_2^2		
X_3	0	0	σ_3^2	
X_4	0	0	0	σ_4^2

Matrix B				
	X_1	X_2	X_3	X_4
X_1	σ_1^2			
X_2	Cov(2,1)	σ_2^2		
X_3	Cov(3,1)	Cov(3,2)	σ_3^2	
X_4	Cov(4,1)	Cov(4,2)	Cov(4,3)	σ_4^2

Matrix C				
	X_1	X_2	X_3	X_4
X_1	σ_1^2			
X_2	Cov(2,1)	σ_2^2		
X_3	0	0	σ_3^2	
X_4	0	0	Cov(4,3)	σ_4^2

diagonals are the covariance terms. Matrix A represents the independence assumption in the completely randomized design. It is clear that all the covariance terms are zero, indicating mutual independence. Matrix B represents exchangeable random variables where all pairs of variables are dependent. Matrix C represents the interaural correlation case of Coren and Hakstian (1990), where the two measures on a particular subject are dependent but independent of the measures of other subjects.

For the general case of exchangeable random variables, Zimmerman et al. (1990) derived the result

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{[SS_x + SS_y][1 + (n-1)\rho_{xx'}]}{n(n-1)(1 - \rho_{xx'})}}}, \quad (1)$$

where SS_x and SS_y are the sums of squares, $n_1 = n_2 = n$, and $\rho_{xx'}$ is the population value of the intragroup correlation, or the population standardized covariance. Equation 1 is a special case of another equation that allowed for intergroup correlation as well as intragroup correlation. It should be noted that in Equation 1 it is assumed that $\rho_{xx'} = \rho_{yy'}$ and that the $\rho(X_1, X_2) = \rho(X_3, X_4) = \rho_{xx'}$. That is, the intragroup correlation is equal for both groups and for all pairs of random variables within a group.³ Because of the assumption of exchangeable random variables, the inflated Type I error rates based on Equation 1 are even more inflated than those found in the interaural correlation case.

For the interaural correlation problem, in which we assume variance-covariance Matrix C of Table 1, the corrected t formula is

$$t' = \frac{t_{\text{obs}}}{\sqrt{\frac{(1 + \rho_{xx'})(n-1)}{n - (1 + \rho_{xx'})}}}, \quad (2)$$

where the notation is the same as for Equation 1. Here t_{obs} denotes the observed, inflated value of t calculated from the data based on dependent observations. This result is derived simply by employing fewer covariance terms in the Zimmerman et al. (1990) derivation in order to model the interaural correlation problem described in Matrix C of Table 1. All that is required to apply the result is to divide the observed t statistic t_{obs} (obtained from analyzing all of the data as if it were independent), by the correction factor indicated. This value of t can then be compared with tabled critical values, as usual.

To apply Equation 2, a researcher needs a sample estimate of $\rho_{xx'}$. This estimate can be obtained directly from the data by calculating a product-moment correlation coefficient (just as in Coren and Hakstian, 1990), or the estimates given in Table 1 of Coren and Hakstian can be used as benchmark estimates of $\rho_{xx'}$. Later in this paper, an example will be given of how to calculate the corrected t statistic.

COMPUTER STUDY

A small monte carlo simulation study was conducted to examine how Type I errors are inflated when measures are dependent, as described above, and whether Equation 2 adequately corrects the inflated Type I error rate. The computer-simulation methodology was similar to that described by Zimmerman et al. (1990). A computer program was written to generate pairs of samples from pseudorandom numbers that were transformed to induce a variance-covariance pattern given in Matrix C of Table 1. Correlations were induced by adding a multiple of one random variable to the other random variable. The multiplicative constant was chosen to obtain the desired correlation value, and the variables were further transformed to have a specified mean and variance. All simulations were based on a normal distribution, obtained by using a Box-Muller (Box & Muller, 1958) transformation on uniformly distributed pseudorandom numbers.

The objectives of the simulation study were to examine the performance of the tests under the null distribution of no difference and under alternative hypotheses of population mean differences. Standardized differences between population means were introduced so that Type I errors, Type II errors, and power could be examined. In this study, the sample sizes were 12, 16, and 20, and there were 5,000 replications of the sampling procedure for each mean difference. Differences between means were 0 to 4.0 δ in incre-

ments of $.5\delta$, where δ is the standard error of the difference between means (see Hays, 1988, p. 304).

An ordinary Student's t test (called the incorrect model by Coren & Hakstian, 1990) was performed on the sample values. Next, the corrected t value was computed using Equation 2, which involves the population correlation, $\rho_{xx'}$. Then the corrected t value was computed from Equation 2, using a sample estimate of $\rho_{xx'}$. Finally, the paired data were averaged and an ordinary Student's t test was conducted on the means, as recommended by Coren and Hakstian. All tests were nondirectional at a significance level of .05.

RESULTS OF SIMULATIONS

Table 2 shows the empirical Type I error rate (the probability of rejecting the null hypothesis given a $\delta = 0$). All of the results in Table 2 were obtained with equal variances and sample sizes, $n = 12$ and $n = 20$. The first column, labeled $\rho_{xx'}$, is the degree of interaural correlation, which varies from 0 to .90 in increments of .10. The column labeled t is the probability value for the ordinary Student's t test assuming the incorrect model. The column labeled t' is the probability obtained using Equation 2 with $\rho_{xx'}$. The column labeled t'' is the probability using Equation 2 with $r_{xx'}$ calculated from the sample. Finally, the column labeled t_{av} is one of the methods suggested by Coren and Hakstian (1990). The column labeled t_{av} is the probability resulting from the ordinary Student's t test calculated on the reduced sample by averaging the two measures on each subject.

The results are consistent with the argument presented by Coren & Hakstian (1990). A correlation of .80 or .90 results in inflated Type I errors ranging from .140 to .175. Furthermore, it can be seen clearly from Table 2 that as $\rho_{xx'}$ increases, the Type I errors increase. It is also clear that all of the alternative procedures correct the probabilities of Type I errors close to the nominal significance level. Finally, comparing the results from t' to t'' makes it clear that using the sample product-moment correla-

n	$\rho_{xx'}$	δ	t	t'	t''	t_{av}
12	.60	0	.119	.051	.053	.050
		.5	.180	.082	.082	.073
		1.0	.336	.185	.177	.158
		1.5	.539	.359	.344	.321
		2.0	.727	.564	.538	.509
		2.5	.868	.740	.716	.692
		3.0	.948	.879	.859	.839
		3.5	.984	.951	.941	.930
		4.0	.997	.984	.979	.975
		4.0	.106	.054	.052	.050
12	.40	.5	.145	.078	.078	.075
		1.0	.256	.163	.156	.151
		1.5	.420	.303	.285	.278
		2.0	.602	.476	.447	.436
		2.5	.769	.653	.617	.601
		3.0	.883	.806	.773	.762
		3.5	.954	.908	.882	.870
		4.0	.985	.963	.950	.943
		4.0	.153	.058	.060	.052
		.5	.235	.107	.110	.095
16	.75	1.0	.443	.264	.257	.239
		1.5	.680	.492	.480	.457
		2.0	.857	.715	.709	.687
		2.5	.962	.884	.875	.860
		3.0	.992	.969	.964	.959
		3.5	.999	.993	.993	.990
		4.0	.999	.999	.999	.998

tion, $r_{xx'}$, in the modified formula is nearly as effective as using the population value, $\rho_{xx'}$.

Table 3 shows the empirical Type I error rate and power at eight values of δ . All results in Table 3 were obtained from the normal distribution with equal variances and sample sizes, $n = 12$ and $n = 16$. The first column, labeled n , is the sample size. The second column, labeled $\rho_{xx'}$, is the degree of correlation. The remaining columns, labeled t , are the values for the ordinary Student's t test assuming the incorrect model (t), using Equation 2 with $\rho_{xx'}$ (t'), using Equation 2 with $r_{xx'}$ calculated from the sample (t''), and finally using means of measures as suggested by Coren and Hakstian (1990) (t_{av}).

As in Table 2, the Type I error rates are maintained at the nominal significance level in all the modified methods. However, the tests based on Equation 2 are more powerful than the one based on averaging the measures from the two ears. (The results would not be different for simply summing across the two ears. In fact, the power and Type I error rates for summing and averaging, or any other linear function of the two-ear measures, will be the same as those in Table 3). Note that it is incorrect to compare the values of $\delta \neq 0$ for t versus the alternative methods because the Type I error rate is not maintained for t .

In summary, the results of the present simulation study are consistent with Coren and Hakstian's (1990) argument and give some indication of the degree of inflation of Type I error rates that can be expected in practice. Furthermore, they indicate that it is possible to effectively correct for these inflated values in research practice. We have

Table 2
Type I Error Rates of the Statistical Tests
as a Function of Intragroup Correlation

$\rho_{xx'}$	$n = 12$				$n = 20$			
	t	t'	t''	t_{av}	t	t'	t''	t_{av}
0	.050	.050	.045	.047	.051	.051	.050	.052
.1	.060	.048	.047	.051	.060	.046	.046	.042
.2	.068	.048	.047	.049	.080	.056	.052	.053
.3	.091	.050	.053	.052	.081	.046	.045	.047
.4	.101	.050	.051	.050	.093	.043	.042	.041
.5	.114	.052	.054	.050	.114	.056	.054	.050
.6	.129	.050	.049	.044	.120	.047	.045	.040
.7	.151	.056	.059	.051	.141	.057	.058	.055
.8	.159	.057	.058	.050	.163	.059	.060	.057
.9	.173	.061	.062	.050	.179	.049	.050	.043

Note— t is the Student's t test performed on the original data; t' is from Equation 2 using $\rho_{xx'}$; t'' is from Equation 2 using $r_{xx'}$; and t_{av} is the t test on the average of the two ears.

demonstrated Coren and Hakstian's point using a general approach that makes it possible to extend these methods to other experimental designs in which nonindependence is a problem. Also, we have demonstrated that our corrected t test is somewhat more powerful than the method that Coren and Hakstian recommend. This is not surprising, since averaging (or summing) across pairs of correlated values substantially reduces degrees of freedom. It is well known that, everything else held constant, an increase in sample size results in an increase in the power of a statistical test. Since the interaural correlations are quite large in magnitude, our method is somewhat akin to taking a copy of the data (n) and conducting the analysis of the entire set ($2n$). The increase in sample size results in a marked increase in power to detect a difference when one exists.

EXAMPLE OF COMPUTATIONS

What follows is an example of fictitious data generated with a $\rho_{xx'} = .80$. The data were generated with the same algorithm used in the previous simulation study. Data were generated from a normal distribution with equal variances and $n_1 = n_2 = n = 12$. The population means were adjusted to a mean of 10 so that the results would appear realistic, and the δ used in 1.50 (a value that could be reasonably found in research settings). Table 4 contains the data for the computational example.

Treating this data with the incorrect model, the resulting statistic is $t(22) = -3.027$, and the critical value of t is 2.074 for a two-tailed test at a significance level of .05. The conclusion is that there is a statistically significant difference between the means of the two groups. However, this decision should be considered within the context of the inflated Type I error probability.

If we were to average or sum across the two measures for one subject, then the test statistic is $t(10) = -2.131$, and the accompanying critical value is 2.228. The conclusion, using the averaging procedure, is that the two groups are not significantly different.

Using Equation 2, we find that

$$t' = \frac{-3.027}{\sqrt{\frac{(1+.80)(11)}{12 - (1+.80)}}} = -2.1726,$$

and we compare this value to the critical value of 2.074 (which is the same critical value as in the incorrect model because of the same degrees of freedom) and conclude that the means are significantly different. Finally, we can compute t' with the sample product-moment correlation coefficient instead of $\rho_{xx'} = .80$. By pairing the data points and calculating the correlation given the 12 pairs, we obtain an $r_{xx'} = .889$. Taking this value and substituting it for $\rho_{xx'}$ in Equation 2, we obtain a value of $t'(22) = -2.1116$. This value, based on a sample correlation,

Table 4
Data for the Computation Example

Group 1			Group 2		
Subject	r.v.		Subject	r.v.	
1	X_1	15.649	7	X_{13}	12.63
1	X_2	15.9	7	X_{14}	12.381
2	X_3	13.654	8	X_{15}	13.686
2	X_4	13.876	8	X_{16}	14.197
3	X_5	15.128	9	X_{17}	15.31
3	X_6	15.457	9	X_{18}	14.474
4	X_7	14.675	10	X_{19}	13.387
4	X_8	15.211	10	X_{20}	12.348
5	X_9	16.175	11	X_{21}	14.281
5	X_{10}	15.735	11	X_{22}	14.701
6	X_{11}	13.707	12	X_{23}	14.15
6	X_{12}	13.967	12	X_{24}	13.805

Note—r.v. is the abbreviation for random variable, and its elements refer to the notation in Figure 1.

also leads to the conclusion that the group means are statistically significantly different. Therefore, for the above example, the Coren and Hakstian (1990) method did not detect the true difference ($\delta = 1.50$), while our more powerful method did.

FURTHER COMMENTS

The purpose of this paper was to provide further evidence to support the conclusions of Coren and Hakstian (1990) and to present some alternative methods of dealing with the interaural correlation issues they raised. Recent developments enable one to use a corrected t test that turns out to be more powerful than a procedure based on averaging subject's scores. The latter solution halves the effective sample size, thereby reducing the degrees of freedom and the probability of detecting a difference when one exists. The method we have suggested is based on a general formula that applies to dependent random variables, Equation 2, and allows the test to be conducted at twice the sample size as can Coren and Hakstian's approach. We conclude that it is not only possible but more desirable to count ears instead of heads as long as one takes the dependence into account and uses a modified t formula. Also, in the context of dependent random variables, the model can be tailored to particular psychonomic research. The issue of nonindependence applies not only to interaural correlation, but also to vision, hemispheric laterality research, and any situation in which two lateral measures are made on one subject. Furthermore, the same issue arises in other experimental contexts in which statistical dependence of other kinds exists.

REFERENCES

- BOX, G. E. P., & MULLER, M. (1958). A note on the generation of normal deviates. *Annals of Mathematical Statistics*, 29, 610-611.
- COREN, S., & HAKSTIAN, A. R. (1990). Methodological implications of interaural correlation: Count heads not ears. *Perception & Psychophysics*, 48, 291-294.

- EDGINGTON, E. S. (1980). *Randomization tests*. New York: Dekker.
- HAYS, W. L. (1988). *Statistics* (4th ed.). New York: Holt, Rinehart & Winston.
- MAGUIRE, T. O., & ROGERS, W. T. (1989). Proposed solutions to non-randomness in educational research. *Canadian Journal of Education*, **14**, 170-181.
- MAY, R. B., MASSON, M. E. J., & HUNTER, M. A. (1989). Randomization tests: Viable alternatives to normal curve tests. *Behavior Research Methods, Instruments, & Computers*, **21**, 482-483.
- WILEY, T. L., OVIATT, D. L., & BLOCK, M. G. (1987). Acoustic-immittance measures in normal ears. *British Journal of Speech & Hearing Research*, **30**, 161-170.
- ZIMMERMAN, D. W., WILLIAMS, R. H., & ZUMBO, B. D. (1990). *Effect of nonindependence of sample observations on parametric and nonparametric statistical tests*. Working Paper 91-9, Edumetrics Research Group, University of Ottawa (manuscript under review).

NOTES

1. To the authors' knowledge, the only noted exceptions of statistical tests that do not make the assumption of independence within groups

are randomization tests. Under nonrandom sampling, the randomization model of inference tests the hypothesis of randomness rather than independence. These tests apply to situations in which we can calculate the exact distribution of the test statistic across all possible permutations of the data at hand. For further details, see Edgington (1980), Maguire and Rogers (1989), and a computer program by May, Masson, and Hunter (1989).

2. Note that individual subject scores can be treated as realizations of a random variable.

3. Given Coren and Hakstian's (1990) results that the between-ear correlations are about the same as the test-retest correlations for one ear, this assumption is not too restrictive. Also, the combined larger sample size can be useful when estimating the within-ear product-moment correlation.

(Manuscript received January 2, 1991;
revision accepted for publication April 2, 1991.)