MAXIMIZATION OF SYSTEMATIC VARIANCE IN RESEARCH DESIGN

Denis Achung Uyanah
Department of Curriculum and Instructional Technology, Cross River University of Technology, Calabar, Nigeria.

ABSTRACT: *The main focus of a research design, whether in pure or applied sciences, is the study of variances in the collected data. Technically, the systematic variance is maximized, the error variance is minimized and the effects of extraneous variables are controlled. In the pure sciences, the maximization of the systematic or desirable variance is done by a good spread in the level of the factors in the study by pulling them apart. In behavioral sciences it is quite easy when the factor(s) are categorical or inanimate. The levels of the factor(s) are deliberately pulled apart. It is a problem when the factor is continuous. Several methods have been advocated. In this study, five of such methods are compared- the use of sample mean and standard deviation, theoretical mean and standard deviation, the correlation coefficients from the transposition of Person by Item Matrix, factor score and factor analysis methods. Two validated instruments designed to measure students attitude towards mathematics and tendency to cheat in examinations, were administered on a sample of 100 students of Cross River University of Technology, Calabar, Nigeria. The students' scores were grouped on basis of their attitude towards Mathematics, using the five methods. One-way ANOVA was carried out with the categorized Mathematics Attitude score as factor and their tendency to cheat in examinations as the dependent variable. The proportions, of the total variance, in the dependent variable, accounted for, in each of the five methods, were compared using the Fishers' Z-test, for all possible pairs of the explained variances. The results showed that four out of the ten paired comparisons were significant, with grouping using sample mean and standard deviation, accounting for the total variance highest. The grouping base on sample mean and standard deviation is recommended and the implications for behavioral research design discussed.*

KEY WORDS: systematic variance, person by item matrix, transposed matrix, factor analysis, factor scores, correlation, factor loadings.

INTRODUCTION

One of the basic principles of research design is that of maximizing the systematic or desirable variance, minimizing the error variance and controlling the effect of the extraneous variables. This principle was referred to MAXMINCON by Kelinger and Pedhazur (1973) and Kerlinger (1986). They agreed that when both dependent and independent variables are continuous, regression analysis should be used in the study of the contribution of the independent variables in the prediction model. However, there are situations in which the researcher is interested in the interaction effect of the independent variables, particularly when some of the independent variables are categorical. In such cases, they recommended that the independent continuous variables may be categorized, to serve this purpose. Their argument was that:

ISSN 2056-3620(Print), ISSN 2056-3639(Online)

1. The interaction with other independent variables may explain a relatively large proportion of the total variances, which the effect of the independent variables considered in isolation and under the assumptions of multiple regression analysis, cannot reveal.

2. The researcher may be afforded greater control and sensitivity of the tests for significance applied. Consequently, they advocated, as a means of reducing the error variance, the identification of, as many sources of systematic variance in the dependent variable, as possible and building such sources into the study, provided such inclusion is feasible and meaningful, under a given set of circumstances.

When an independent variable is continuously measured as recommended by Nunnally (1978), using a validated instrument, the data so generated closely approximates true continuous measure, having the qualities of ratio leveled measured variable (Kelinger, 1986). This makes the carrying out of any arithmetical process, valid. So that descriptive parametric and inferential analysis can be applied. Such statistics include, but not limited to the computation of means, standard deviation, correlation coefficients, and factor analysis. In situations where the continuous variable is measured using the Likert scale or a modification of it, the sample mean and standard deviation can be computed, normally, as with other continuous random variables.
On the basis of algebra of expectation and summations, Frank and Althoen (1994) have shown that the theoretical mean can be computed as $\mu^1=n\mu$ where n=number of scaled items and $\mu=\frac{1}{2}$ (a+b) and variance $\sigma^2 \frac{n^2}{2}(a+b)^2$where "a" represents the minimum score on the scale and (b) is the maximum. From this, the theoretical standard deviation can be computed.

A correlation matrix obtained from an item-by-person matrix is a measure of inter-item correlation. When the item-by-person matrix is transposed i.e. rows become columns and columns become rows-the correlation matrix is a collection of inter-person relationships. The interpretation of such correlation coefficients is that positive correlation means the persons are always associated with each other. Similarly a negative correlation means, the two persons are always moving in opposite directions such that when one increasing in the measured variable the other is decreasing. This correlation coefficient may be significant or not. This means a group of positive correlations is a group of persons identical on the basis of the attribute measured by the items or variable. The implication is that groups of persons can be identified using their measured relationships.

When Factor Analysis is done using the transposed matrix, literally, the factor loadings represent how they click to each other to form a group. The number of factors obtained using principal component analysis validly represents the number of such groups identifiable in the sample. All that is needed is that the sample should be heterogeneous enough, an image of the population from which the sample was randomly selectedWhen factor analysis is carried out on the original item by person matrix, to the full, by obtaining factor scores, the persons can be grouped on the basis of their factor scores, using the mean and standard deviation of their individual factor scores.A one-way ANOVA of the dependent variable can be carried out with the categorized independent variable as a factor. This may be logically extended to two or three or more-Way ANOVA. The between-groups-variance is the variance of the dependent accounted for by the factor(s), and the

ISSN 2056-3620(Print), ISSN 2056-3639(Online)

ratio of the explained variance to the total variance is the proportion of the total variance in the dependent variable, accounted for by the factor(s). These are the theoretical basis of this study. These concepts and procedures have been tried out and found to be worth further investigation by Uyanah (2011 & 2014). The results were quite encouraging with respect to categorization using sample mean and standard deviation, theoretical mean and standard deviation and inter-person correlation when item-by-person matrix is transposed.

METHODOLOGY

Two validated instruments, designed to measure student Mathematics attitude and their tendency to cheat in examination, whose measure of internal consistency obtained using the Cronbach Alpha were 0.786 and 0.814, were administered on a random sample of 100 students of the Cross River University of Technology, Calabar, Nigeria. The independent variable was the students' attitude towards Mathematics and the dependent variable was their tendency to cheat in examinations. The respondents' scores on the dependent variable were grouped on the basis of their score (X) on the independent variable.

Method 1: Using the sample mean ($\overline{X}$) and standard deviation (s)

Grouped:

1. $X > \overline{X} + 3S$
2. $\overline{X} + 2S \leq X \leq \overline{X} + 3S$
3. $\overline{X} + S \leq X < \overline{X} + 2S$
4. $\overline{X} - S \leq X \leq \overline{X} + S$
5. $\overline{X} \_ 2S < X < \overline{X} - S$
6. $\overline{X} - 3S < X \leq \overline{X} - 2S$
7. $X \leq \overline{X} - 3S$

Method 2: Using the same criteria as in method 1 but using the theoretical mean (μ) and standard deviation (σ).

Method 3: Using correlation coefficient with a positioned hypothetical person

Group 1: persons whose correlation with the positioned person was negative and not significant.

Group 2: Those whose correlation coefficient with the positioned person, was negative and significant.

Group 3: Those whose correlation coefficient with the positioned person, was positive and significant.

Group 4: Those whose correlation coefficient with the positioned person, was positive but not significant.

Method 4: Grouping based on factor loadings.

Factor analysis, using the principal component approach was done on the transposed item-by-person matrix. Seven factors were extracted and those whose factor loading on a particular factor was greater than .200 made that group.

Method 5: Factor analysis with Factor scores

ISSN 2056-3620(Print), ISSN 2056-3639(Online)

Full factor analysis was carried out on the item-by-person matrix and obtained factor scores for each respondent. The method 1 was now repeated but with the means and standard deviation of the factor scores replacing the sample mean and standard deviation.

One-way ANOVA of the respondent's tendency to cheat in examinations was carried out using each of the factors obtained from method 1 to 5 as independent variable. The proportion of the total variance in their tendency to cheat was computed to obtain R_1 to R_5. The Rs were then compared using Fishers Z transformation method and in pairs.

RESULTS AND INTERPRETATION

The descriptive statistics that preceded each of the ANOVA are present for each method in Table 1

Table 1

Descriptive statistics of tendency to cheat: Methods by Groups

Method	Group	N	Mean	Std. Dev.	Std. Error
Sample mean and standard deviation	Above 91	10	25.00	5.000	1.581
	Between 63.5 and 91	35	47.71	2.688	.454
	Between 36.2 and 62.5	40	50.38	4.317	.683
	36 and below	15	49.67	12.454	3.216
	Total	100	46.80	13.996	1.400
Theoretical mean and standard deviation	Above 98.9	10	25.00	7.071	2.236
	Between 70 and 98.9	20	47.75	8.261	1.847
	Between 41.1 and 69	55	49.64	10.764	1.451
	Below 41.1	15	49.67	21.572	5.570
	Total	100	46.80	13.996	1.400
Using Correlation Coefficients	Negative Insignificant. r	7	26.30	8.135	3.075
	Negative significant r	43	27.52	9.026	1.376
	Positive significant r	45	30.69	11.116	1.657
	Positive insignificant r	5	42.86	14.268	6.381
	Total	100	46.80	13.996	1.400
Factor analysis with transposed matrix	Factor 1	30	44.00	13.538	2.472
	Factor 2	20	53.25	8.139	1.820
	Factor 3	20	40.50	21.548	4.818
	Factor 4,5,6, or 7	30	49.50	9.834	1.795
	Total	100	46.80	13.996	1.400
Factor analysis with factor scores	Above 2.0	20	46.00	4.163	.931
	Between 0.0 &1.9	30	39.67	14.814	2.705
	Between -1.9 & 0.0	30	49.33	13.095	2.391
	Below -1.9	20	54.50	16.543	3.699
	Total	100	46.80	13.996	1.400

ISSN 2056-3620(Print), ISSN 2056-3639(Online)

These descriptive statistics were not compared because that was not the focus of the study. They may just be noted, in case of a replication of this study.
The results of the one-way ANOVA of tendency to cheat, by groups for each of the grouping methods are presented in Table 2

Table 2
One-way ANOVA of tendency to cheat in examination by categorized attitude towards Mathematics

Method of categorizing	Source of variation	Sum of square	df	Mean square	F-value	R-value	P-value
Sample mean and Std. Dev.	Between Group	1083.230	3	361.077	14.890*	.318	.002
	Within Group	2327.970	96				
	Total	3411.200	99				
Theoretical mean and Std. Dev.	Between Group	1067.238	3	355.746	14.570*	.313	.003
	Within Group	2343.962	96	24.416			
	Total	3411.200	99				
Inter-Person Correlation	Between Group	50.000	3	16.667	.475	.015	.689
	Within Group	3361.200	96	35.103			
	Total	3411.200	99				
Factor analysis on X_T	Between Group	415.950	3	138.650	4.444	.122	.004
	Within Group	2995.250	96	31.201			
	Total	3411.200	99				
Factor scores mean & Std. Dev.	Between Group	583.533	3	194.511	6.605*	.171	.003
	Within Group	2827.667	96	29.455			
	Total	3411.200	99				

Significant at .05 level. P<.05.

The results in Table 2 show that the proportion of the total variance accounted for by the factor was highest when grouping was done using sample mean and standard deviation (R^2 = .318), followed by that when theoretical mean and standard deviation were used (R^2 =.313) and the least was the R^2 obtained using the inter-person correlation coefficients (R^2 = .15).
The obtained R-square values were compared using Fishers Z-transformation technique for all possible pairs. The results are in table 3

ISSN 2056-3620(Print), ISSN 2056-3639(Online)

Table 3
Fishers' Z-test comparison of the proportion of explained variance in tendency to cheat in examination

Explained variance	R^2_1	R^2_2	R^2_3	R^2_4	R^2_5
R^2_1	.318**	.439	4.181*	1.907	1.381
R^2_2	.005	.313**	4.137*	1.863	1.337
R^2_3	.316	.311	.015**	2.273*	2.799
R^2_4	.196	.191	.120	.122**	.526
R^2_5	.147	.142	.169	.049	.171

Significant at .05 level. $Z_{Cal.} > Z_{Crit.} = 1.96$
values along main diagonal are proportions of explained variance, above it are calculated Z-values and below it are corresponding difference in transformed Z-values.

From Table 3, the differences in explained variable between methods 1 and 3 (Z_{Cal} =4.181), methods 2 and 3 (Z_{Cal} =4.137), methods 3 and 4 (Z_{Cal} =2.273) and methods 3 and 5 (Z_{Cal} =2.799) are significant, because the calculated Z-values are greater than the critical Z-value (1.96). All the other paired comparisons were not significant.

DISCUSSION OF FINDINGS

In this study, five methods of grouping independent variables for the purpose of applying ANOVA were compared. The motivation was drawn from the increasing desire of researchers, especially in behavioral sciences, to categorize independent variables and apply ANOVA as well as from the relative advantage that factorial studies have over other multivariate statistics, observed by Kerlinger (1986) and Kerlinger and Pedhazur (1973).

The results have shown those methods of categorizing, that significantly superior to others advocated in theory. That they did not all lead to the same decision, indicated that they are not equivalent. What is even more interesting is the fact that when categorization was done using the sample mean and standard deviation, the proportion of the total variance in the dependent variable, accounted for by the factor in ANOVA, was relatively higher than all others.
These results are in the same direction with those of Uyanah (2014) and Uyanah (2011), giving a support to the first three methods, though in this study, categorizing and inter-person correlations led to a different decision.

The method that one anticipated would be superior was the factor analysis method, because of the orthogonal relationship between factors. This needs further investigation, though they seem to support the argument made by Kerlinger (1986), leaning a lot of weight on the validation of the instruments than the sample.It should however be noted that the categorization advocated here over-rides the disadvantages associated with it as enumerated by Kerlinger (1986) and Kerlinger and Pedhazur (1973). These disadvantages may be undermined only in extremely necessary settings.

ISSN 2056-3620(Print), ISSN 2056-3639(Online)

References

Frank, H. & Althoen, C.S.(1994). *Statistical concepts and applications.* London: Cambridge University Press

Kerlinger, F.N. (1986). *Foundation of Behavioural research* (3rd ed.). New York: Holt, Rinehart and Wilston

Kerlinger, F.N. & Pedhazur, E.J. (1973). *Multiple regression in behavioural research.* New York: Holt, Renehart and Wilston.

Nunnally,(1978).*Psychometric theory*.

Uyanah, D. A. (2011). *Desirable variance maximization in research design*. A paper presented at Faculty of Education, University of Calabar, in Partial fulfillment of the requirement for the award of Ph.D in research and statistics.

Uyanah, D.A. (2014). *Desirable variance maximization in research design.* A Paper Presented at the 16th Annual Conference of Association of Educational Researchers and Evaluators of Nigeria, 14-20th July 2014, International Conference Centre, University of Calabar, Nigeria.

ISSN 2056-3620(Print), ISSN 2056-3639(Online)