MASS MEDIA USE AND POLITICAL KNOWLEDGE: A MODEL FOR MASS COMMUNICATIONS RESEARCH USING ONE-WAY ANALYSIS OF VARIANCE PROCEDURES

Sayed Hussain Pasha
Department of Information
King Saud University
Riyadh

ABSTRACT

This article is an attempt to outline a model of one-way analysis of variance for mass communications research. The article first presents the ANOVA model showing how the different calculations are carried out. And then, the model uses some fictitious data to illustrate this model. The data have been arranged so as to make calculations easy and yield easily interpretable results. It is hoped the model would be of pedagogic value. The hypothesis generated in the model could be tested with a proper sample with what may turn out to be quite useful and interesting results.

This article is an attempt to outline a model for the use of Analysis of Variance procedures in mass communications research. ANOVA is a statistical procedure for assessing differences in group means.[1] Even though best suited to handle experimental or manipulative data, it is flexible enough for use in survey, content and other data with which a mass communications scholar could be expected to deal.

The use of ANOVA is indicated in cases where the dependent variable is measured on a metric or interval scale in terms of real numbers, and the independent variable or factor is measured on a nominal scale in terms of discrete categories or levels. ANOVA could be one- or n-way depending on the number of factors in the model (Kerlinger, 1973: 214-239; Winkler and Hays, 1975:747-767).

The model presented in this article consists of one dependent variable, Political Knowledge, measured on an interval scale and one independent variable, Media Use, measured on a nominal scale. The model, therefore, uses one-way ANOVA. The entire model, however, is hypothetical as the data used in the model are not real. The numbers have been manipulated so as to keep calculations simple, and draw easily interpretable results.[2]

Alman and White (1968:144-156) discuss a number of statistical methods used in mass communications research, including Analysis of Variance procedures. But their approach is more theoretically-oriented than the one adopted in this article. As a result, the Alman and White article is likely to help the readers gain a better insight into the "Why" of doing the various steps involved in the analysis of variance, whereas the present article is expected to help them gain a better understanding of the "How" of those steps.

DEFINITIONS

Definitions of concepts and variables are key elements in any model. Definitions are of many kinds such as Nominal, Real (in the sense of meaning or empirical ananlysis), Verbal, etc. The model presented here uses two kinds of definitions: Conceptual and Operational. Conceptual Definitions mainly provide a verbal picture or description of what a concept or variable means. Operational Definitions, on the other hand, specify the procedures by following which a given concept or variable could be located in the real world. In this sense, Operational Definitions are not definitions at all but merely instructions for measurement (Mullins, 1971:7-35; Hempel, 1972:2-20; Hage, 1972:62-85; Nachnias and Nachnias, 1976:17-20).

The model uses the following definitions:

Concept Name: Political Knowledge
Conceptual Definition: The extent to which people know about what is going on in the world in terms of political and public events.
Operational Definition: Score on a 15-item Knowledge Index whose items would be selected with a view to tap information about local, national and international events figuring prominently in the local news media in the recent past.
Concept Name: Media Use.

Conceptual Definition: The extent to which people make use of the mass media for informational purposes.
Operational Definition: Amount of actual use made of the three media, Newspaper, Radio and Television for informational purposes as measured in response to the question: Generally speaking, when you want to find out what is going on in the world, which of the three media do you use more? Newspaper, Radio or Television?[3]

HYPOTHESIS

If research is the attempt to find solutions to certain scientific problems, the scientific problem itself could be expressed in the form of a question. In this sense, research is really an attempt to locate in the real world answers to certain clearly stated questions of scientific interest.

A hypothesis is a tentative--probabilistic--answer to the research question. It is generally stated in the form of a relationship between two or more variables, or, strictly speaking, between two or more operational definitions. A hypothesis guides the scientist's research efforts, and is the central focus of data generation, analysis and interpretation.

Research Questions:
Generally speaking, mass media are purveyors of information, and a good bit of media attention is focused on political and public affairs information. It is, therefore, reasonable to expect that the use of the mass media would be a strong correlate of political and public affairs knowledge. While Newspaper, Radio and Television, all carry political and public affairs information, it is probably the Newspaper which **as a medium** carries more information of this kind than either one of the other two media. Therefore, the Newspaper could be expected to be a more effective medium where political and public affairs information was concerned than either Radio or Television. The research question for the purposes of the present model would, then, be as follows:

> Do the three newsmedia, namely, Newspaper, Radio and Television differ significantly in terms of their impact on political and public affairs knowledge?[4]

Hypothesis:
A hypothesis, as noted earlier, is a tentative answer to the research question. It generally takes the form of a stated relationship between two or more variables. The tentative answer to the question, 'If the three media (Newspaper, Radio and Television) affect politial and

public affairs knowledge differntially', is that they probably would. Here, the two variables of interest are Media Use and Knowledge. Normally, a hypothesis asserting a relation between these two variables would state that Media Use would have a significant positive correlation with Knowledge. But since the focus of the present model is on seeing if and how the three media Newspaper, Radio and Television differ among themselves in their impact on Knowledge, and since ANOVA is a procedure for testing the group means, the research hypothesis would turn out to be as follows:

> The Knowledge means for the three media, Newspaper, Radio and Television, would be significantly different from one medium to another.

This hypothesis would be tested against the Null Model which would state that in terms of population parameters, there is no difference among the three means, and any observed differences in the sample means would be due purely to chance (Meyer, 1973:64-96). Thus.

$$H_0 = \mu_{TV} = \mu_{Radio} = \mu_{Newspaper} = \mu$$

or, alternatively stated,

$$H_0 = \mu_{TV} - \mu_{Radio} - \mu_{Newspaper} = O.$$

This simply means that in terms of population parameters, the three means are not different. As against this, the research hypothesis or H_1 would be,

H_1 = Not H_0

This would mean,

$$H_1 \neq \mu_{TV} \neq \mu_{Radio} \neq \mu_{Newspaper}$$

Or alternatively stated,

$$H_1 = \mu_{TV} - \mu_{Radio} - \mu_{Newspaper} \neq O$$

This simply means that the three means come from different populations, and the observed difference/s among sample means are not entirely attributable to chance factors.

THE DATA

Data for testing this hypothesis could be obtained by asking a sample of 30 young people from the local university or college to fill out a self-administered questionnaire constructed to tap the two variables, Media Use and Knowledge.[5] A sample of 30 is far from adequate for purposes of most serious scientific investigations, but for illustrative purposes, this sample would suffice. The sample could be divided into three equal-sized groups of 10 each, representing the three media categories considered in the model. The data obtained from the questionnaire could, then, be analyzed using ANOVA procedures.

ANOVA: SOME BASIC IDEAS

Essentially, variance is a measure of spread, or dispersion of the scores around the mean which is a measure of central tendency. The data must have variation if research is to be meaningful. It is variation in the dependent variable, Y, which is of primary interest in research. Where there is no variation in the data, everything would point to the mean, and there would really be nothing to investigate. The scientist's chief interest lies in decomposing the total variability of Y into its component parts and tracing each segment to its source such as a clearly identifiable independent variable or random fluctuations untraceable to a specific known factor.

Variance estimates are obtained by calculating the deviations of scores about the mean, squaring them, and then dividing the sum of squared deviations by N - 1, that is by the total number of scores less 1. The formula for such a calculation would be,

$$v = \frac{\Sigma x^2}{N - 1}$$

where V is variance, Σx^2 is the sum of squared deviations, and N - 1 is the total number of scores minus 1. The numerator in the formula, Σx^2, is referred to as the "Sum of Squares".

Analysis of Variance is essentially a process of calculating sums of squares, attributing them to their sources and determining if any observed differences among them are statistically significant, that is, if they could have arisen purely due to chance factors. In one-way ANOVA, there are three sets of Sums of Squares, namely, Sum of Squares Total (SST), Sum of Squares Between Groups (SSB) and Sum of Squares Within Groups (SSW). SST provides an indication of the toal variability in the dependent variable, Y. This includes variation due to the independent variable as well as due to error. SSB is indicative of the effect of the independent variable on the dependent variable, and SSW is indicative of the error variance, that is, random fluctuation of scores within each group.

Thus,

$$SST = SSB + SSW$$

and

$$SSW = SST - SSB.$$

Analysis of Variance could be calculated entirely with raw scores. In that event, however, the formula must include a correction for the mean by using the Correction Factor. The use of the Correction Factor in effect reduces raw scores to deviation scores. The CF, or Correction Factor, is obtained by summing all the raw scores in all the groups, squaring the sum and then dividing the squared sum by N, that is by the total number of scores. Thus, the raw score formula for calculating SST will be as follows:

$$SST = \Sigma X_T{}^2 - \frac{(\Sigma X T)^2}{N}$$

where $\Sigma X_T{}^2$ is the sum of raw scores for all groups each individually squared, $(\Sigma X_T)^2$ is the sum total of all the raw scores first summed and then squared.

SSB is calculated by summing and squaring the raw scores belonging to each group separately, dividing the squared sum for each group by the number of observations in that group, n j, adding all the quotients together and then finally subtracting the Correction Factor from their sum. The formula for SSB then would be as follows:

$$SSB = \left[\frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \ldots + \frac{(\Sigma X\, p)^2}{n\, p} \right] - CF$$

where $\frac{(\Sigma X)^2}{n}$ is the sum of the raw scores for each group squared and divided by the number of scores in that group. The subscripts indicate the group to which the calculations belong. CF stands for Correction Factor.

SSW as indicated earlier is calculated by subtracting SSB from SST.

Sums of Squares are divided by their respective degrees of freedom (df) to obtain Mean Squares. The degrees of freedom for SSB are equal to the number of groups less 1 (p - 1). The degrees of freedom for SSW are equal to the total number or scores less the number of groups (N - p). Thus, Mean Square Between Groups (MSB) is obtained by dividing SSB by p - 1, and Mean Squares Within Groups (MSW) is obtained by dividing SSW by N - p. Finally, MSB is divided by MSW to get the F value which is checked in the table of F distribution against the critical value to see if the obtained value of F is significant at a given level of probability.

All these calculations are then thrown into a summary table which looks like Table 1.

TABLE 1

Analysis Of Variance Summary Table (Dummy)

Source of Variation	Sums of Squares	Degrees of freedom	Mean Squares	F	Level of Significance of F	E
Between Groups	SSB	p - 1*	SSB/p-1	MSB/MSW	P** or α	$\sqrt{SSB/SST}$
Within Groups	SSW	N - p	SSW/N-p			
Total	SST	N - 1				

* p in this column refers to the number of groups in the model
** p here refers to probability of a finding resulting from chance.

RESULTS

Table 2 provides the raw data for the hypothetical sample of 30 young persons from the local university or college. The table consists of six columns, two for each one of the three media categories considered in the model. The first column in each category provides the raw Knowledge scores for that category, (X). In the second column, these scores have all been squared, (X^2). Column frequencies have been artificially held constant at 10 for each medium-- a

TABLE 2

Knowledge scores for the three media categories

Newspaper		Radio		Television	
X	X^2	X	X^2	X	X^2
13	169	6	36	10	100
10	100	6	36	6	36
9	81	7	49	6	36
7	49	8	64	6	36
12	144	10	100	9	81
11	121	4	16	9	81
10	100	5	25	4	16
10	100	5	25	5	25
9	81	7	49	5	25
8	64	10	100	7	49
$\Sigma X_1 = 99$	$\Sigma X_1^2 = 1009$	$\Sigma X_2 = 500$	$\Sigma X_2^2 = 500$	$\Sigma X_3 = 67$	$\Sigma X_3^2 = 485$

convenient arrangement which does not often occur in real life research situations. The last row in Table 2 provides the totals for Knowledge scores for each media category, (Σx and Σx^2).

Looking at Table 2, it is not difficult to arrive at the conclusion that the Knowledge scores are considerably higher for the Newspaper group than for the other two groups. Column 1 in the Newspaper group has six double digit entries whereas the same column for the Radio and Television groups has only two and one respectively. These differences in individual scores are reflected in the totals of columns 1 and 2 in each category.

These impressions are confirmed by examining Table 3 which provides information about the number of cases as well as the means and standard deviations for the three groups.

TABLE 3

Number of cases, means and standard deviations

Media Groups	Number of cases	Means	Standard Deviations
Newspaper	$n_1 = 10$	$\bar{X}_1 = 9.9$	$SD_1 = 1.70$
Radio	$n_2 = 10$	$\bar{X}_2 = 6.8$	$SD_2 = 3.76$
Television	$n_3 = 10$	$\bar{X}_3 = 6.7$	$SD_3 = 3.61$
All three groups	$N = 30$	$\bar{X}\,T = 7.8$	$SD\,T = 2.06$

According to Table 3, the mean for the Newspaper group (9.9.) is considerably higher than the mean for either the Radio or Television group which are respectively 6.8 and 6.7.

Table 4 provides the summary statistics required for the various ANOVA calculations. Section A of Table 4 provides the totals for column 1 for each of the three media categories in Tabl 2, ΣX. In section B of Table 4, the same totals have been squared, $(\Sigma X)^2$. Section C of Table 4 provides the sums of squared scores, ΣX^2, that is, the totals for column 2 for each one of the three media categories in Table 2.

TABLE 4

Sums required for ANOVA calculations

A. ΣX = Sums of scores from column 1 in each of the three media categories in Table 2

Newspaper Group	ΣX_1	99	
Radio Group	ΣX_2	68	
Television Group	ΣX_3	67	
All three groups	$\Sigma X\,T$		234

B. $(\Sigma X)^2$ = Sums of scores from column 1 in each of the three media categories in Table 2, squared

Newspaper Group	$(\Sigma X_1)^2$	9809	
Radio Group	$(\Sigma X_2)^2$	4624	
Television Group	$(\Sigma X_3)^3$	4489	
All three groups	$(\Sigma X\,T)^2$		54756

C. ΣX^2 = Sums of scores from column 2 in each of the three media categories in Table 2

Newspaper Group	ΣX_1^2	1009
Radio Group	ΣX_2^2	500
Television Group	ΣX_3^2	485
All three groups	$\Sigma X\,T^2$	1994

Table 5 provides the various ANOVA calculations such as the sums of squares, mean squares, etc. Section A of Tabe 5 gives the Correction Factor which must be subtracted from the calculations for various sums of squares as correction for the mean. Sections B and D of Table 5 show how the sums of squares and mean squares are calculated. Section C provides the df needed for calculating mean squares, while section E provides the F value and the way of calculating it.

TABLE 5

ANOVA Calculations

A. Correction Factor:

$$\mathbf{CF} = \frac{(\Sigma X\,T)^2}{N} = \frac{54756}{30} = 1825$$

B. Sums of Squares:

$$SST = \Sigma X_T^{\,2} - CF = 1994 - 1825 = 169$$

$$SSB = \left[\frac{(\Sigma X)^2}{N_1} + \frac{(\Sigma X_2)^2}{N_2} + \frac{(\Sigma X_3)^2}{N_3} \right] - CF$$

$$= 1890 - 1825 \qquad = \; 65$$

$$SSW = SST - SSB = 169 - 65 = 104$$

C. Degress of Freedom:

$SSB = p-1 = 3-1 = 2$

$SSW = N - p = 30-3 = 27$

D. Mean Squares:

$$MSB = \frac{SSB}{p-1} = \frac{65}{2} = 32.5$$

$$MSW = \frac{SSW}{N-P} = \frac{104}{27} = 3.85$$

E. F

$$\frac{MSB}{MSW} = \frac{32.5}{3.85} = 8.44$$

Table 6 provides the summary of the ANOVA calculations carried out in Table 5. Table 6 is an exact replica of Table 1 with one important difference: Table 1 was a dummy set up to work as a model whereas Table 6 contains actual numbers in terms of the data from Table 2. According to Table 6, the SSB and SSW are 65 and 104 respectively. These, when divided by their respective degrees of freedom (2 for SSb and 27 for SSW), yield MSB = 32.5 and MSW = 3.85. Dividing MSB by MSW (32.5/3.85), an F value of 8.44 is obtained which according to the F Table is significant beyond the .01 level.

F is a ratio (MSB:MSW) which provides an indication of the amount of variance accounted for in the data by the factors, as a proportion of error variance, that is, variance which remains unexplained.[6] The probability or alpha level indicates the number of times in a hundred, a thousand or ten thousand cases that an outcome could be a result of pure chance. In the present instance, there is one chance in a hundred ($\alpha = .01$) that an F value so high could have been obtained purely due to the element of chance operating in the data.

TABLE 6

Analysis of Variance Summary Table

Source of Variation	Sums of Squares	Degrees of freedom	Mean Squares	F	Level of significance of F	E
Between Groups	65	2	32.50			
			8.44	$< .01$	.61	
Within Groups	104	27	3.85			

The last column in Table 6 provides the E statistic which is a correlation ratio between the dependent and independent variables. Unlike F, the E statistic is a measure of the strength of relationship between two variables. This statistic is obtained by first dividing SSB by SST and then taking out the square root of the quotient. When squared, E^2, like R^2, provides an indication of the proportion of variation in the dependent variable accounted for by the independent variable.

DISCUSSION

The results show that Media Use significantly impacts Knowledge. An E^2 of .37 indicates that 37 percent of the variation in Knowledge could be accounted for by Media Use. An F value of 8.44, which is significant beyond .01 level, indicates that the three media categories differ significantly among themselves in terms of their impact on the dependent variable Knowledge. But how exactly do they differ, that is, which category differs significantly from which category? The answer to a question like this is not directly provided by ANOVA. To get a question like this answered, certain other tests, mostly of an a-posteriori nature, are indicated. Some such tests are Newman-Keuls Test, Tukey's HSD (honestly significant differences) Test and Scheffe's Test.

In the present model, however, no such tests were used, as they were not considered necessary. The results are clear-cut and unambiguous. Radio and Television means, which are 6.8 and 6.7 respectively, are almost identical. This means, there is no difference between these two media groups in terms of impact on Knowledge. In comparison to this, the Newspaper mean is very much different from the other two. the Newspaper group has a mean of 9.9 as compared to the joint mean of 6.75 for the Radio and Television groups combined, a difference of 3.15. This is a sizeable difference, and judging from the F value, a significant one. This means, the entire impact of Media Use on Knowledge is coming from Newspaper. The significant F implies a rejection of the null hypothesis of No Difference (H_0), and constitutes research support for the alternative hypothesis, H_1. [7]

CONCLUSION

The model presented in this article, illustrated the use of One-Way ANOVA procedures for investigating a postulated bivariate relationship between Media Use and Political and Public Affairs Knowledge. Using hypothetical data, the model demonstrated calculation techniques for the various sums of squares and mean squares needed for ANOVA purposes. The F-ratio was calculated to indicate the amount of variance explained in the data as a proportion of error (unexplained) variance. The level of significance of the F value was obtained to assess the extent of chance operating in the data. The E statistic and E^2 were obtained as measures of the strength of the bivariate relationship.

In line with the research question and hypothesis presented in the model, data were arranged to make one of the three means--for the Newspaper group--come out significantly different from the other two. But since the model was mainly designed for illustrative and methodological purposes, no substantive implications drawn from the findings could be considered valid.

NOTES

1. Use of ANOVA is indicated in cases where the number of groups to be compared, exceeds 2. Otherwise, in cases where a direct comparison between two means is intended, the appropriate statistical technique to use is the t-Test.

2. With large data sets, hand calculations are nearly impossible. As a result, the researcher is forced to rely on computers for the analysis of large data basis. While knowledge of a computer language such as FORTRAN is a definite asset, commonly available computer package programs like SPSS (Nie et al. 1975:398-433) and SAS (Helwig and Council, 1979) are adequate for most research purposes related to the mass media.

3. Roper (1975:3-10) has been using a similar question for a longitudinal study of mass media preferences in the United States. The exact question wording used in this study was: "First, I would like to ask you where you usually get most of your news about what is going on in the world today--from the newspapers or radio or television or magazines or talking to people or where?" (Roper, 1975:3). These surveys revealed that during a period between 1959-1974, the newspapers declined as the most popular news source in America (from 57% in 1959 to 47% in 1974), while television registered a consistent gain during the same period (from 51% in 1959 to 65% in 1974).

The newspapers, however, have not been without their supporters, and the Roper findings have been challenged by other researchers in the field. Carter and Greenberg (1965:29-34) found that newspaper— and not television--was the source most widely depended upon for news by most Americans. Forty-four percent of the sample used by Carter and Greenberg named newspapers as their most depended upon news source as compared to 32 percent naming television. Findings by Clarke and Ruggels (1970:464-471) support a similar conclusion.

4. Here the question of interest is not media preference as in the case of some of the studies cited earlier (cf. Roper, 1975), but the actual relationship of the use of a given medium such as newspaper, radio or television, to resultant political and public affairs knowledge.

5. Since most persons carrying out research of this kind are expected to be associated with a college or university, a student sample is probably one of the most easily accessible ones for their purposes. It is also likely to be the least expensive one. Social scientists, including communication scholars, have made extensive use of convenient student samples even though such use is not without problems (Pasha, 1982:265-274)

6. Generally speaking, variance is of two kinds: Systematic variance and error variance. Systematic variance is that portion of variation in the data which is due to one or more clearly identified independent variables. This type of variance is known by different names in the literature such as Between group variance, Variance due to treatment, Variance due to factor, Explained variance, etc. Error variance does not have a specific identifiable source. It is a product of random and often self-compensating fluctuations and influences within the data. It includes variance due to some likely independent variables unknown to the researcher. It is that portion of variation in the data that is not accounted for by the independent variables in the model. This type of variance is known in the literature as Unexplained variance, Within group variance, Residual variance, etc.

7. The all important question before the researcher is: To be or not to be--that is, to reject or not to reject the null hypothesis, H_0 should be rejected only when it is false and not when it is true. By rejecting the H_0 when it is true, the researcher commits one type of error---the Type I Error. By failing to reject the H_0 when it is false, the researcher commits another type of error----the Type II Error. In the first instance, he mistakenly infers the existence of a significant relationship between

two variables where there is actually none. In the second instance, he does the opposite: he fails to take note of a relationship that actually exists. The two errors are so intertwined that reducing the probability of making one, automatically increases the probability of making the other. A researcher, therefore, is required to make a conscious decision as to the type and amount of risk he would be willing to take.

Successfully rejecting the null hypothesis, however, does not automatically "prove" that the alternative hypothesis (the research hypothesis) is the correct one. All it does is to render the alternative hypothesis a little more plausible than it was before. That is, out of a whirling mess of influences working on a variable, the researcher has successfully eliminated one false possibility. By doing this, he has edged that much closer to a correct mapping of reality. But how much more he has to go still, is a question which is by no means easy to answer in non-experimental research which relies primarily on statistical controls as a means of isolating causal influences.

REFERENCES

ALMAN, J. and D. WHITE, "Statistical Methods in Communication Research", in R. Nafziger and D. White, **Introduction to Communication Research.** Baton Rouge: Louisiana State University Press, 1968.

CARTER, R., and L. RUGGELS. "Preferences among newsmedia for coverage of public affairs", **Journalism Quarterly,** 47, Autumn: 1970.

HAGE, G. **Techniques and Problems of Theory Construction in Sociology.** New York: John Wiley and Sons, 1972.

HELWIG, H., and COUNCIL, K. (eds.) **SAS User's Guide,** Carry, N.C: SAS Institute, 1979.

HEMPEL, C. **Fundamentals of Concept Formation in Empirical Science.** Chicago: Chicago University Press, 1972.

KERLINGER, F. **Foundations of Behavioral Research.** New York: Holt, Rinehart and Winston, 1973.

MEYER, P. **Precision Journalism.** London: Indiana University Press, 1973.

MULLINS, N.**The Art of Theory.** New York: Harper and Row, 1971.

NACHNIAS, D., and C. NACHNIAS. **Research Methods in Social Sciences.** London: Edward Arnold, 1976.

NIE, N., HULL, C., JENKINS, J., STEINBRENNER, K., and BENT, D. **Statistical Package for the Social Sciences,** New York: Mc-Graw-Hill, 1975.

PASHA, S. "Social Science Literature: Perspectives and Problems", **Arab Journal for the Humanities,** 2:7 (summer, 1982), pp. 265-274.

ROPER, B. **Trends in Public Attitudes Toward Television and Other Mass Media 1959-1974.** New York: Television Information Office, 1975.

WINKLER, R., and W. HAYS, **Statistics.** New York: Holt, Rinehart Winston, 1975.