MUTUAL INFORMATION APPROACH TO ANALYSIS OF ENTROPY IN ONE AND TWO-WAY DISCRETE STATISTICAL DESIGNS

A.T. Sóyínká[1], A.A. Olósundé[2], F.S. Apantaku[1], A.O. Wálé-Òròjo[1]

[1] Federal University of Agriculture - Alabata, Abeokuta, Ogun state - Nigeria, Department of Statistics
[2] Obafemi Awolowo University - Ile-ife, Osun state - Nigeria, Department of Mathematics

Corresponding author: Ajíbólá Táíwò Sóyínká, soyinkaat@funaab.edu.ng

ABSTRACT: The dynamics of research in terms of well-defined study design, interwoven plots/sampling frames, research stages, sample proportions in each plot, nature of output(s) and sample size for different designs can be extensively studied using the theory of mutual information. Owing to the fact that outcomes in multi-dimensional contingency platform are frequentist in nature and as such the usual analysis of variance is not practicable; there is need to develop a sound mathematical techniques for entropy's $H(.)$ significance among interacting vectors whose subsets are discrete and are interwoven into plots in a categorical set up. This study partitions the mutual information algebraic structures via the squared radial function of the exponential power distribution into sum of single and joint entropy's shared contributions in analogy to the analysis of variance sum of squares partitions. We present evidence-based test of entropy significance in few statistical designs with application.

KEYWORDS: Discrete multi-dimensional design, mutual information, covariance matrix, algebraic structure, entropy significance

1. INTRODUCTION

Analysis of variance and covariances are important decision-making tools that are commonly used in research and in any discipline where statistical decision theory is the bed-rock of empirical evidence. Unfortunately, despite the popularity of analysis of variance (ANOVA), its restriction to data on the continuous scale is a major drawback that must be addressed. Hence, we developed a technique that will be useful in hypothesis testing of data that are best captured on a nominal/ordinal scale via categorical set up. The interactions among the discretized variable(s), which resulted into several interwoven plots, are jointly combined into a multi-dimensional categorical set up. The joint probability outcome across several plots within the entire categorical set up are presented in frequency form as the number of samples within each plot [22, 24]. Extending the approach to several dependent vectors, the measure of dependence among vector's and its test of significance, across the entire multi-dimensional categorical outline, for every single and joint vector were obtained using mutual information approach; which generalizes the entropy measure. Thus, this study established the mutual information measures (entropy significance) of the relationship among several dependent but discrete vectors arranged into one- and two-way statistical design, but follows multivariate exponential power distribution density. Unlike [7] that used generalized likelihood ratio of several independent log-linear models to capture the same relationship. On the other hand, despite the importance of sample size in research, many of the sample size estimations in literature are not without gaps that only sound statistical theories can fill. Those gaps include not able to accommodate researchers subjective reasoning in design to effect research dynamism [8], usage of arbitrary participants to variable ratio for soft landing on sample size [10], assumption of unrealistic limiting theorem in sample size calculation; most especially when a more practical distribution was not employed in obtaining the limiting theorem [15], and so on. So, in order to address the issues, ANOVA impracticability and provision of sample size that is design dependent, first we established the mutual information analogous of analysis of variance (ANOVA), then extends its usefulness, via a more practical distribution that generalized normal and many other distributions, to the estimation of sample size in some statistical designs; while creating an opening to extend the technique to sample size in more complex designs.

According to [11], the random variable rv X is univariate exponential power distributed (EPD) if

$$f_X(x)=\frac{1}{\sigma\Gamma(1+\frac{1}{2\beta})2^{1+\frac{1}{2\beta}}}e^{-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}};\quad 1.1$$

$$-\infty<|x-\mu|<\infty,\sigma>0,\beta>0$$

Where β is the shape parameter, μ *and* σ are location and scale parameter. Also, its multivariate extension for $[X_1,X_2...,X_p]$ in the absolute continuous case has the density

$$f_{X_1,X_2,...,X_p}(x_1,x_2,...,x_p)=|\Sigma|^{-(1/2)}\frac{\Gamma(1+\frac{p}{2})}{\pi^{\frac{p}{2}}\Gamma(1+\frac{p}{2\beta})2^{1+\frac{p}{2\beta}}}\quad 1.2$$

$$\exp^{-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu)^\beta}$$

Where $E[X_1,X_2..,X_p]=[\mu_1,\mu_2,..,\mu_p]$ and

$$Var[X_1,X_2..,X_p]=\left[\frac{2^{\frac{1}{\beta}}\Gamma[\frac{p+2}{2\beta}]}{p\Gamma[\frac{p}{2\beta}]}\Sigma\right]$$ with

covariance Σ [5, 20]. Note that the squared radial function which is the exponent in (1.2) is dependent on the varying in-built shape parameter β within the density generator function h^p [2, 3, 5,]

$$s\nexists(X-\mu)^T\Sigma^{-1}(X-\mu)|^\beta\sim\Re^2(h^p)\ 1.3$$

However, by varying β in (1.3), the square radial function of the various members of the EPD will be obtained. The squared radial function obtained when $\beta=1$ is our focus in this study.

On the other hand, employing the survival function approach to obtaining discrete probability model [23], the discretized cdf and pdf for (1.1) can be obtained as

$$F(x)=-\left(\frac{q^x}{p}\right)^\beta\ln\left(\frac{q^x}{p}\right)^\beta\quad 1.4$$

and

$$f(x)=F(x)\left(q^\beta\left[\frac{(x+1)\ln q-\ln p}{x\ln q-\ln p}\right]-1\right),\quad 1.5$$

$x=0,1,2,..,0<(p,q)<1.$

Proof. Substituting $\frac{\mu}{\sigma}=\ln p^{-1}$ and $\frac{x}{\sigma}=\ln q^{-x}$ into the series representation of the cdf of (1.1), we have (1.4). Evaluating $f(x)=F(x+1)-F(x)$ we have (1.5).

From (1.5), evaluating the expression $H(x)=-\sum_x\left[\frac{f(x)}{F(x)}\ln\left(\frac{f(x)}{F(x)}\right)\right]$, we obtained the univariate entropy measure for discretized EPD over x as

$$H(x)=\sum_x\left(q^\beta\left[\frac{(x+1)\ln q-\ln p}{x\ln q-\ln p}\right]-1\right)\ln\left(q^\beta\left[\frac{(x+1)\ln q-\ln p}{x\ln q-\ln p}\right]-1\right)\quad 1.6$$

The multivariate discretized EPD is thus expected to generalize the entropy measure (1.6) via mutual information approach. Next we established the relationship between the covariance structure (1.3) in continuous case and the mutual information measure of the entropy (1.6) in discrete case.

2. RELATING MULTIVARIATE MUTUAL INFORMATION (MMI) AND COVARIANCE STRUCTURES OF p^{th}VECTORS

2.1. Algebraic model structure of MMI.

Definition 1: The mutual information (MI) between two random vectors $(rv's)$ X and Y is

$$I(X,Y)=H(X)+H(Y)-H(X,Y)\quad 2.1$$

Definition 2: [21] The MMI of three $(rv's)$ X,Y,Z is

$$I(X,Y,Z)=H(X)+H(Y)+H(Z)-H(X,Y)-H(X,Z)-H(Y,Z)+H(X,Y,Z)\quad 2.2$$

Definition 3: The conditional multivariate mutual information (CMMI) [2, 21] for dependent vectors is

$$I(X_1,..,X_{p-1}|X_p)=I(X_1,..,X_{p-1})-I(X_1,..,X_{p-1},X_p)\quad 2.3$$

Implying that for one input (X) and one output (Y) we have

$$I(X|Y)=I(X)-I(X,Y)=H(X,Y)-H(Y)$$

while two inputs and one output MI has the structure $I(X,Y|Z)=H(X,Z)+H(Y,Z)-H(Z)-H(X,Y,Z)$.

Definition 4: The MMI structure for two inputs and two outputs is [6, 13, 21]

$$I(X,Y|Z,A)=H(X,Z,A)+H(Y,Z,A)-H(Z,A)-H(X,Y,Z,A)\quad 2.4$$

2.2. Relationship between Covariance matrix and MMI

Definition 5: Given two random vectors (X,Y), the covariance matrix Σ is given as

$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{pmatrix}$. Likewise the MI structure for vectors (X,Y) from definition (1) can be expressed in matrix form as

$$\exp^{I(X,Y)} = \begin{pmatrix} \exp^{H(X)-\frac{1}{2}H(X,Y)} & 0 \\ 0 & \exp^{H(Y)-\frac{1}{2}H(X,Y)} \end{pmatrix}.$$

Definition 6: Given three random vectors (X,Y,Z), the covariance matrix Σ is given as

$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} & \Sigma_{xz} \\ \Sigma_{xy} & \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{xz} & \Sigma_{yz} & \Sigma_{zz} \end{pmatrix}$, while its corresponding MMI matrix structure from definition (2) is

$$e^{I(X,Y,Z)} = \begin{pmatrix} \exp^{H(X)-H(Y,Z)+\frac{1}{3}H(X,Y,Z)} & 0 & 0 \\ 0 & \exp^{H(Y)-H(X,Z)+\frac{1}{3}H(X,Y,Z)} & 0 \\ 0 & 0 & \exp^{H(Z)-H(X,Y)+\frac{1}{3}H(X,Y,Z)} \end{pmatrix}.$$

Hence following the univariate

$I(X) = -\frac{1}{2}\ln(1-\rho^2)$ [14] and the multivariate

$$I(X_1, X_2, .., X_n) = A\left(\frac{1}{2}\ln|\Sigma| + \left[\ln D + \frac{1}{\beta}\right]\right)$$

[18] dependence model between mutual information and covariance structure which considers correlation as a measure of linear dependence, we derived the linear relationship between MMI and covariance structure has [2, 3, 14]

$$\exp^{I(X_1, X_2 ..., X_p)} \propto \Sigma. \quad 2.5$$

Substituting (2.5) into (1.3), we expressed the squared radial function of the EPD for p^{th} vectors $X = (X_1, X_2, .., X_p)$ in terms of MMI entropy's as

$$s = \left|(X-\mu)\exp\left(-I\left[X_1, X_2, ..., X_p\right]\right)(X-\mu)\right|^{\beta} \sim \Re^2(h^p). \quad 2.6$$

So for multivariate normally distributed pth vectors $(\beta = 1)$ we have

$$s = \left|(X-\mu)\exp\left(-I\left[X_1, X_2, .., X_p\right]\right)(X-\mu)\right| \sim \chi^2_p. \quad 2.7$$

3. SAMPLE SIZE ESTIMATION IN DISCRETE STATISTICAL DESIGN

Supposing the values from every vector $X_1, X_2, ..., X_p$ are interwoven to map out an area that is common to all the vectors $X_1 \cap X_2 \cap .,.,. \cap X_p$, then any point within the mapped area has coordinate values from each vector. If the point is the means coordinate $\bar{X}$, then its confidence interval for the parameter estimates μ from (2.7) can be obtained as

$$Pr\left(\bar{X} - \sqrt{\chi^2_{p,\frac{1-\alpha}{2}} e^{I(X_1, X_2, ..., X_p)}} < \mu < \bar{X} + \sqrt{\chi^2_{p,\frac{1+\alpha}{2}} e^{I(X_1, X_2, ..., X_p)}}\right) = 1-\alpha. \quad 3.1$$

See [9] for the proof. (3.1) implies that the interwoven vectors discrete points that form the individual plots/ sampling frames within the entire multi-dimensional contingency platform are distributed around a location parameter μ and confined within a probability confidence interval of $Pr\left(\left|(X-\mu)\exp\left(-I\left[X_1, X_2, ..., X_p\right]\right)(X-\mu)\right| \le \chi^2_p\right) = 1-\alpha.$

However, note that in a multi-dimensional contingency table, the sample size (n) is the total sum of the relative frequency (f_r) across all interwoven plots $n = \sum f_r = 1$; hence the sample size (n) required to carry out a research study where the variable (treatments/interventions/vectors /causes/pattern/effects and so on) are jointly interacting and simultaneously influencing each other within the multi-dimensional contingency platform can be derived as

$$n = \frac{\chi^2_{p,\alpha} e^{I(X_1, X_2, ..., X_p)}}{|(X-\mu)^T (X-\mu)|}. \quad 3.2$$

Proof. See [17] for details. Recall that the sample size (n) for the univariate case $(p=1)$ of EPD is given as

$$n = \frac{[tan^{-1}(Z_{\frac{\alpha}{2}} + Z_{\lambda})]^2}{|\bar{X} - \mu|^2}\sigma^2, 0 < \beta < 1 \quad 3.3$$

Or

$$n = \frac{[e^{(Z_{\frac{\alpha}{2}} + Z_{\lambda})} - 1]^2}{|\bar{X} - \mu|^2}\sigma^2, \beta \ge 1, \quad 3.4$$

where σ and μ are as earlier defined, $\bar{x}$ is the univariate sample means or its multivariate means coordinate $\bar{X}$, $Z_{\frac{\alpha}{2}}$ and Z_{λ} are the standard normal values at specified level of significance α and power $(1-\lambda)$ respectively, then considering (2.5), (3.3), (3.4) and noting that σ^2 is equivalent to Σ in the multivariate case while the square of p^{th} $rv'z'$ is $\chi^2_{p,\alpha}$ distributed, then the sample size for the discrete version EPD in a multi-dimensional contingency experiment represented as

$$n = \frac{[tan^{-1}(Z_{\frac{\alpha}{2}} + Z_{\lambda})]^2 e^{I(X_1, X_2, ..., X_p)}}{|(X-\mu)^T (X-\mu)|}, 0 < \beta < 1 \quad 3.5$$

$$n = \frac{[e^{(Z_{\frac{\alpha}{2}} + Z_{\lambda})} - 1]^2 e^{I(X_1, X_2, ..., X_p)}}{|(X-\mu)^T (X-\mu)|}, \beta \geq 1. \quad 3.6$$

can be expressed as (3.2). Note: In practice, the reference sample point co-ordinate may be taken from previous studies or estimated, since the true mean co-ordinate is usually unknown.

4. ANALOGY BETWEEN SUM OF SQUARES IN CONTINUOUS CASE AND SUM OF ENTROPY'S IN DISCRETE CASE

Taking the natural logarithm of (2.7); we obtain the sums of each entropy components in the MMI algebraic structure in analogy to the sum of squares in continuous case analysis of variance (ANOVA) as

$$|I[X_1, X_2, ..., X_p]| \sim \ln[(X-\mu)^T (\chi_p^2)^{-1} (X-\mu)] \quad 4.1$$

Note: (4.1) showed that, the algebraic sum of entropies is the logarithm of the ratio of two chi square. Recall that ratio of two $\chi's$ is Fisher-distributed. So from (4.1), provided $0 < |I_b - I_a| < \infty$ since MI must be non-negative; then we deduced that the distribution of MMI difference $|I_b - I_a|$ is analogically similar to the F-distribution for the right hand index (4.1) ratio $\frac{MS_{Regression}}{MS_{error}} \sim F \geq 1$ usually used in ANOVA. Hence, the MMI algebraic structure has the distribution

$$\exp(-|I_b - I_a|) \sim 0 < F < 1. \quad 4.2$$

which is the inverse of existing F - distribution. Note that (4.2) also follows from the discretization of the continuous Fisher $(\frac{m}{2}, \frac{n}{2})$ distribution via the series expansion of its incomplete beta cdf. Next, we discuss the MMI algebraic structure of few statistical designs to demonstrate the usefulness of entropy significance and ascertain its level of dependency (odds ratio).

5. STATISTICAL DESIGNS

5.1. One-way MI design

In this case, we have a single treatment random variable X (p=1) which could be in various subsets; that is partitioned into two or more categories of treatment effects $X = [X_1, X_2, ..., X_z]$ such that each categories has its own level of measurement or factor levels for example

$$X = X_1[-1,1], X_2[1,2,3], ..$$

$$., X_z[\text{Low}, \text{Medium}, \text{High}, \text{Severe}].$$

Supposing Y is the outcomes across all the X's treatment effects then the one-way design has the MI structure $I(X,Y)$. For single trial $H(Y) = 0$, the conditional MI is $I(X|Y) = H(X) - H(X,Y)$ while for multiple outcomes the MI is $I(X,Y) = H(X) + H(Y) - H(X,Y)$. Example is set of psychiatric instruments with different scales aimed at screening patients for psychosis. While the continous case one-way ANOVA answers the hypotheses: average screening effectiveness across the instruments (treatments) are the same or not, the MI approach, answers the hypothesis : the entropy of agreement on effectiveness among the instruments are the same all through or not; that is $H_o : I(X|Y) = 0$ and $H_a : I(X|Y) \neq 0$ while for multiple trials we have $H_o : I(X|Y) = 0$ and $H_a : I(X,Y) \neq 0$. The entropy that accounted for the significance in the alternative hypothesis will then be determined via MI analysis of entropy significance in analogous to the ANOVA [1, 4]. The preliminary research design structure on how treatments are arranged into plots is given as

$$\begin{pmatrix} x & y & \dots & z \\ -1 & 1 & \dots & Low \\ 1 & 2 & \dots & Medium \\ \dots & 3 & \dots & High \\ \dots & \dots & \dots & Severe \end{pmatrix}$$

making the least preliminary number of plots in the design to have a row matrix $X' = [2,3,...4]$.

Table 1: One-way MI design analysis without replication

source	Estimate (I)	Df	$\lvert I_{trt} - I_{error} \rvert$	$\lvert F_{cal}$	$F^{-1}_{tab, df_{I_{trt}}, df_{I_{error}}, \alpha}$
$H(X)$	I_{trt}	r	.	.	.
$H(X,Y)$			.	.	.
$-H(X)$	I_{error}	n-r			
$H(X,Y)$	I_{total}	n			

Trt : Treatment

5.2. Two-way MMI design

In this case, we have two treatment random variables [X,Y] (p=2) where each X and Y are vectors separately partitioned into categories such that each category has various factor levels say for instance

$X = x_1[-1,1], x_2[1,2,3], .x_c[\text{Low}, \text{Medium}, \text{High}, \text{Severe}]$ and $Y = y_1[0,1,2], y_2[-1,1], .., y_r[1,2,3,4,5]$

So the design matrix is a rectangular or square array of interacting input vectors (X,Y) with variable output Z and thus has the MI structure $I(X,Y,Z)$. The conditional MI for two inputs given an output is $I(X,Y \mid Z) = H(X,Z) + H(Y,Z) - H(Z) - H(X,Y,Z)$ where for single trial $H(Z) = 0$. Example is investigating the effect of specific drug therapy dosages (factor X) and occupational therapy interventions (Factor Y) among patients with third psychiatric episodes in-terms of length of stay on admission as output (Z). The aim is to determine if drug therapy, occupational therapy and its interactions has the same or significantly different entropy's contributory effect $(H_0 : I(X, Y \mid Z) = 0$ and $H_a : I(X,Y \mid Z) \neq 0)$ [4, 9]. The matrix of how treatments are arranged to interact with each other and the number of interwoven plots available in the design is the Cartesian products of interacting categories given as

	$x_1[-1,1]$	$x_2[1,2,3]$	...	$x_c[Low, Medium, High, Severe]$
$y_1[0,1,2]$	6	9	...	12
$y_2[-1,1]$	4	6	...	8
..	.	.	...	.
..	.	.	..	.
..	.	...	.	
$y_r[1,2,3,4,5]$	10	15	..	20

Next, we proceed to determine the entropy that accounted for the significance of the alternative hypothesis.

Table 2: Two-way MI design analysis without replication

source	Estimate (I)	Df	$\lvert I_{trt} - I_{error} \rvert$	F_{cal}	$F^{-1}_{tab, df_{I_{trt}}, I_{error}, \alpha}$
H(X,Z)	I_{trt1}	p	.	.	.
H(Y,Z)	I_{trt2}	q	.	.	.
H(X,Y,Z)-H(X,Z)-H(Y,Z)	I_{error}	pq-p-q	.	.	.
H(X,Y,Z)	I_{total}	pq			

Trt : Treatment.

Note that the MMI algebraic structure of three interacting vectors given the output variable A is $I(X,Y,Z \mid A) = H(X,A) + H(Y,A) + H(Z,A) - H(X,Z,A) - H(X,Y,A) - H(Y,Z,A) + H(X,Y,Z,A)$.

The MMI for other research designs, like r^k factorial designs, split plots, multivariate output designs and so on, can be obtained in a similar fashion.

6. APPLICATION

6.1. Sample Size Estimation.

In this case, assumptions may be taken such that the number of plots in each interwoven categories are assumed to follow a known probability distribution. We demonstrate application using examples.

Example 1: What is the sample size required to carry out an experiment on the effectiveness of some selected psychiatric screening instruments with least preliminary design matrix $X' = [2,3,4]$ when $(\alpha, 1-\lambda) = 0.9$.

Solution: First we assume the uniform distribution and secondly the poisson distribution; so we obtain the probability values across plots $p(X') = [1/3, 1/3, 1/3]$ and $p(X') = [4/11, 4/11, 3/11]$ respectively. The sample size (assuming unit deviation) is thus

$$n = \frac{[\arctan(1.64 + 1.28)]^2}{\exp^{\ln 1/3}} = 15163.678 \ for \ 0 < \beta < 1$$

and $n = \dfrac{[\exp^{(1.64+1.28)} - 1]^2}{\exp^{\ln 1/3}} = 923.09 \ for \ \beta \geq 1$. On the other hand for the poisson distribution the sample size is $n = 15034.547$ and $n = 915.23$ for $0 < \beta < 1$ and $\beta \geq 1$ respectively.

Example 2: Supposing the least number of plots in each sampling frame generated from the interactions of two vectors (X,Y) is given by the array $\begin{pmatrix} & x_1 & x_2 & x_3 \\ y_1 & 6 & 9 & 12 \\ y_2 & 4 & 6 & 8 \\ y_3 & 10 & 15 & 20 \end{pmatrix}$. Calculate the sample size to carry out a study design with above description at $\alpha = 0.95$ and $1 - \lambda = 0.9$ assuming unit deviation.

Solution: For assumed uniform and poisson distributions we obtain

$$p(X') = \begin{pmatrix} & x_1 & x_2 & x_3 \\ y_1 & 1/9 & 1/9 & 1/9 \\ y_2 & 1/9 & 1/9 & 1/9 \\ y_3 & 1/9 & 1/9 & 1/9 \end{pmatrix}; and$$

$$p(X') = \begin{pmatrix} & x_1 & x_2 & x_3 \\ y_1 & 0.1 & 0.2 & 0.145 \\ y_2 & 0.03 & 0.1 & 0.172 \\ y_3 & 0.2 & 0.05 & 0.03 \end{pmatrix}.$$

The sample size is

$$n = \begin{pmatrix} & Uniform & Poisson \\ 0 < \beta < 1 & 5306.77 & 6037.58 \\ \beta \geq 1 & 601.90 & 684.79 \end{pmatrix}.$$

6.2. Mutual Information (MI) analysis of entropy significance.

Example 3: Supposing the distribution of respondents screened for anxiety (X), depression (Y) and general health status (Z), is given below. Estimate the associations within and among the screening scales and obtain its effect.

Table 3: Distribution of respondents screened for Depression, Anxiety and GHQ [18]

X	Z	Y_-	Y_+	Total
X_-	$GHQ < 5$	37 32	8 13	45 45
X_+	$GHQ \geq 5$	2 10	3 21	5 31
	Total	81	45	126

Table 4: Three-way MMI design Analysis of entropy Significance

source	Estimate (I)	df	$\lvert I_{trt} - I_{error} \rvert$	F_{cal}	$F^{-1}_{tab,df_{I_{trt},I_{error}},\alpha}$
-H(X,A)	-0.68829	2	3.113407	0.0444	0.19455
-H(Y,A)	-0.501195	2	2.92631	0.05359	0.19455
-H(Z,A)	-0.256426	2	2.681541	0.068	0.19455
H(X,Y,A)	1.004612	4	1.420503	0.24159	0.22075
H(X,Z,A)	0.90356	4	1.52155	0.21837	0.22075
H(Y,Z,A)	0.742952	4	1.682163	0.18597	0.22075
I(X,Y,Z|A)	2.425115	6	0		0.233645
H(X,Y,Z,A)	1.2199	8	1.205215	0.29966	0.240964

Critical region: $F_{cal} < F^{-1}_{tab,df_{I_{trt},I_{error}},\alpha}$.

Interpretation: Each of the instruments are independently significant alongside its joint associations $p-value < \alpha$ except the association between Depression and Anxiety. The odd ratio effect of having GHQ induced anxiety is 1.174 times higher than that of GHQ induced depression with 21.8% *and* 18.6% of the respondents presently showing signs of GHQ induced anxiety and GHQ induced depression respectively.

Example 4: What is the level of dependence/association between current and previous religious Identification [1] in the table below

Table 5:

	Y(CRI)	Protestant	Catholic	Jewish	Others	Total
X(PRI)	Protestant	918	27	1	70	1016
	Catholic	30	351	0	37	418
	Jewish	1	1	28	1	31
	Others	29	5	0	25	59
	Total	978	384	29	133	1524

Solution:

Table 6: MI Analysis of Entropy Significance

source	Estimate (I)	df	$\lvert I_{trt} - I_{error} \rvert$	F_{cal}	$F^{-1}_{tab,df_{I_{trt},I_{error}},\alpha}$
H(X,Z)	0.830233	4	1.29598	0.273629	0.26042
H(Y,Z)	0.9196	4	1.385343	0.250238	0.26042
H(X,Y)-H(X)-H(Y)	-0.46575	8	0	1	0.290698
H(X,Y,Z)	1.28408	16	1.749829	0.173804	0.31056

Significant if $F_{cal} < F^{-1}_{tab,df_{I_{trt},I_{error}},\alpha}$

Interpretation: The joint entropy and the entropy due to CRI are significant at $\alpha = 0.05$. Implying that the entropy between CRI and PRI are signicantly different at $\alpha = 0.05$ and so has a level of dependence. The result shows that only 5.11% of the respondents have changed their religious identification from what it was previously.

6.3. Performance Evaluation of Screening Instruments.

Having established the importance of F-calculated in the confirmation of the significance of entropy's in mutual information (MI) analysis, we proceed to demonstrating its usefulness in the evaluation of screening instruments performance based on the level of mutual agreement of subjects' responses across all the instruments. Given that F-calculated $0 < F_{cal} < 1$ can be used as a single measure of performance of the test accuracy for the positive class [12], then, from table 4 in example 3, we have False Positive (FP) $= \alpha = 0.0444$ while its corresponding True Positive (TP) $= 1 - \alpha = 0.9556$ for the anxiety case; others follow suit in the same order. So based on this argument, we construct the table for the diagnostic classification among the three screening instruments using the F-calculated values as follow:

Table 7. Sensitivity and Specificity of the screening instruments

True value	Instruments	Test value$_{Positive}$	Test value$_{Negative}$
Positive	Anxiety	0.9556T P	0.24159F N
	Depression	0.94641T P	0.21837F N
	GHQ	0.932T P	0.18597F N
Negative	Anxiety	0.0444F P	0.75841T N
	Depression	0.05359F P	0.78163T N
	GHQ	0.068F P	0.8143T N

TP-True Positive; FP-False Positive; FN-False Negative; TN-True Negative

Next, we obtain the various measures of instruments performance from table 7. See [12, 16] for the formula on various measures of classification.

Table 8: Measures of performance among Anxiety, Depression and GHQ Instruments

Measure	Anxiety	Depression	GHQ
Sensitivity	83.71	81.25	79.41
Specificity	94..83	93.58	91.77
Positive Predictive	95.56	94.64	93.2
Value Negative	81.43	78.16	75.84
Predictive Value	16.19	12.66	9.65
Likelihood Ratio	0.1718	0.2	0.2243
Positive Likelihood	88.5	86.40	84.52
Ratio Negative	89.24	87.44	85.76
Accuracy	89.27	87.42	85.59
F-score	0.7854	0.7484	0.7118
Area Under Curve	2.506	2.286	2.074
Youdens Index			
Discriminant Power			

CONCLUSIONS

The major advantage of the theory of MI in research designs is its ability to accommodate design dynamism in terms of the nature and relationship between inputs and output(s) vectors along with the various interaction modifications. Though the algebraic structure may be complex depending on the design under study, the truth however is that the structure and its test of entropy's significance is achievable. Further research in this area of our discussion will open up a future of promising professional discipline for statisticians in all facets of life where research and development is of topmost importance from discrete outcomes. This study established the analogical relationship between discrete case mutual information (MI) entropy significance and continuous case sum of squares significance in ANOVA. The level of dependence via entropy's significance and its odds effect among interacting vectors (variable) were confirmed alongside the performance evaluation.

REFERENCES

[1] **A. Agresti** – Categorical Data Analysis. A Wiley Interscience Publication.ISBN 0-471-85301-1, 1990.

[2] **R.B. Arellano-valley,J.E. Contreras-reyes and G.G. Marc** – Shannon Entropy and mutual information for multivariate and skew-elliptical distribution. *Scandinavian Journal of statistics Theory and applications*, 81: 42-61, 2012.

[3] **D. Brillinger and A. Guha** – Mutual Information in the frequency domain. *Journal of Statistical Planning and Inference,* 137(3):1076-1084, 2006.

[4] **K. Gomez and A. Gomez** – Statistical Procedure Design and Experiment for Agricultural Research. 2nd edition Wiley Interscience Publication, 1984.

[5] **E. Gomez, E., M.A. Gomez-Villegas, and J.M. Martin** – A multivariate generalization of the exponential power family of distributions. *Communications in Statistics*, A27:589-600, 1998.

[6] **T.S. Han** – Multiple Mutual Information and Multiple Interactions in Frequency Data, *Information and Control*, 46(1):26-45, 1980.

[7] **C.S. Hong and B.J. Kim** – Mutual information and redundancy for categorical data. Stat Papers 52, 17–31, 2011.

[8] **J. Charan and T. Biswas** – How to Calculate Sample Size for Different Study Designs in Medical Research. *Indian Journal of Psychological Medicine*, 35(2):121-126, 2013.

[9] **R.A. Johnson and D.W. Wichern** – *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 2006.

[10] **N. Kock and P. Hadaya** – Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. Information Systems Journal 28(1): 227-261, 2018.

[11] **J.K. Lindsey** – Multivariate Elliptically Contoured Distributions for Repeated Measurements. *Biometrics*, 55:1277-1280, 1999.

[12] **P. David** – Evaluation from precision, Recall and F-measure to ROC Informedness, Markedness and correlation. *Journal of Machine Learning Technology*, 2(1): 37-63, 2011.

[13] **C.E. Shannon** – A mathematical theory of communication. *Bell sys. Tech. Journal*, 27(3):379-423, 1948.

[14] **R. Smith** – A mutual information approach to calculating non linearity. The ISI *journal for rapid dissemination of statistical research*. ArXiv: 1512.0075Ovl, 2015.

[15] **U. Sekaran and R. Bougie** – Research methods for business: A skill-building approach (5th ed.). Haddington: John Wiley and Sons, 2010.

[16] **M. Sokolava, N.Japkowicz, S. Szpakowicz** – Beyond Accuracy, F-score and ROC. A Family of Discriminant measures. Australian Joint Conference on Artificial Intelligence 4304:1015-1021, 2006.

[17] **A.T Soyinka** – Generalizing Sample Size of Normally Distributed Samples Using Generalized Exponential Power Distribution. *Journal of the Nigerian Statistical Association*. 28:74-78, 2016.

[18] **A.T. Soyinka, A. A. Olosunde, O.E. Asiribo** – On Mutual Information for Elliptical Distributions: A Case of Nonlinear dependence of 'n' vectors. *Sri Lankan Journal of Applied Statistics (SLJAS)*, 18(1):19-34, 2017.

[19] **T. Soyinka, A. Olosunde, S. Oyetola, A. Akinhanmi** – Modelling Job Satisfaction and Age Using a Composite Function of Gamma and Power Transformation Function, *International Journal of Statistics and Applications*, 7(2) (2017), 107-112.

[20] **A.T. Soyinka and A.A. Olosunde** –Inferences from Asymmetric Multivariate Exponential Power Distribution. Sankhya B Journal of Indian Statistical Institute. Volume 82(1) Advance online publication DOI 10.1007/s13571-020-00235-w 07-08-2020.

[21] **S. Sunil**–*Review on Multivariate Mutual Information*.www.nd.edu/tutorial/sunil, 2003.

[22] **J. Seok and Y. Seon-Kang** – Mutual Information between Discrete Variables with Many Categories using Recursive Adaptive Partitioning. Sci Rep 5, 10981, 2015.

[23] **S. Chakraborty** – Generating discrete analogues of continuous probability distributions-A survey of methods and constructions 2:6: 1-30, 2015.

[24] **O.A. Wale-Orojo A.T. Soyinka, F. S. Apantaku, O.D. Atanda and A.A. Akintunde** – Modeling categorical data in frequency domain via mutual information approach. Anale. Seria Informatică. 18: 64-68, 2020.