Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 4, June 2021:988- 1003

Research Article

Enhancement of Survey Research: A Statistical Perspective

KDV Prasad
Adjunct Professor, Institute of Industrial and Human Resource Development, Financial District, Gachibowli 500 032, Telangana, India. e-mail: prasadkanaka2003@yahoo.co.in

Budumuru Muralidhar
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)
Patancheru 502324, Telangana, India E-mail: b.muralidhar@cgiar.org

Abstract
Surveys are carried out to study the characteristics of a population by studying a sample a sub-set of the population. An accurate sample which is representative of the population is essential to gather representative data to draw statistical inferences and conclusion for decision making using the outcome of the survey. However, there is a lacuna on part of a selection of the type study variables— continuous, nominal, ordinal, or ratio and on the needed statistical test which can be run on the gathered data. This situation leads to wrong representation results and improper decision-making. Therefore, in this review, the researchers elaborate on the selection of appropriate statistical tests for analyzing the survey data and interpretation of results. This manuscript covers the most commonly used statistical stets for survey data association, reliability, predictions, differences between groups, and calculation of sample size for a study. The researchers also provided on the use of binomial, ordinal, and multinomial logistic regressions and the authors assume that researchers carrying out a study is well-versed with sampling methods, and emphasizing the sampling methods is out of the scope of this paper. The researchers believe that the students, research scholars, and other research communities who undertake the survey research will immensely benefit from our article.

Keywords: Survey, *continuous, ratio, variable, ordinal logistic regression, lacuna, population*

Introduction

Surveys are generally carried out to study the characteristics of a population using a representative sample, which is a subset of the whole population. The representativeness is the main characteristic of a good sample and which is turn a miniature or small portion of the whole sample. It is the job of a researcher who is carrying out survey research/study to see that sample the main and focal characters of the sample gender, age, behavior patterns, health issues are represented systematically in the sample, for example, if a study population consists 10,000 people, 50% are female, and 40% who are over 60 years of age, then a representative sample consists fewer people (may be 1000). Also consists 50 of female 40% over the age of 60. It is necessary for the survey researchers to obtain a sample that is representative of the target population of his/her study.

The sample can only be considered based on its accuracy with which characterizes the target population which embodies systems, organization, issues, and to identify a solution to a problem and research survey results are applied and generalized. The survey will have a purpose, specific objectives, and research questions related to the problem for which the survey being carried out. In general, the researchers will define the group/gender/criterion to include and exclude from the sample, and based on this research questions will be developed. Once the survey objective, inclusion, and exclusion criterion are developed, a researcher will select appropriate sampling methods from probability sampling and nonprobability sampling. Probability sampling will lay the statistical foundation to prove that the selected sample is the representative/miniature of the study or target the whole population. Further, in probability sampling, every unit/member of the population will have a fair chance to be included in the sample. In non-probability sampling, the respondents will be selected based on the judgment of a surveyor toward the characteristics of a target population. Therefore, there is a fair chance of biases Fink, (2013). Simple random sampling, stratified random sampling, systematic sampling, cluster sampling, multistage sampling ARE the most commonly used random sampling methods. Convenience sampling, snowball sampling, quota sampling, focus groups are the most commonly used non-probability sampling methods.

Once the sampling selection is completed and data gathered sampling errors and non-sampling errors will be identified and the errors due to the circumstances which are outside the sampling process. The non-sampling errors are common and mostly due to ambiguous definitions of the study population and errors in the design of the survey instrument and measurement. After gathering the data, most of the researchers will assess the normality of the data distribution. Normally this is done based on the estimation of standard error and standard deviation. For normally distributed data, the range is 6 times of standard deviation (Andre Framcis, 2008) sample mean is estimated and this means will be used along with standard deviations measuring several factors like low, medium, and high values; for example, low-performance effect, high-performance effect and moderate performance effect (Sumathi and Nandagopal, 2014).

After estimating the appropriate sample size, confidence levels will be established based on the binomial characteristics using statistical tables and these tables will be used only when errors appear due to sampling and for random sampling. Before estimation of sampling size, certain factors need to be considered. The most common consideration is a grouping of all the survey objectives, research questions, clearly defined unambiguous hypotheses—the syntax is objective, question, and hypothesis. Each of the hypotheses consists of independent or predictor variations and dependent or outcome variables. The authors assume the researchers have perfect knowledge defining the independent, dependent variables, and further to divided/categorize these variables, identification f subgroups, data collection needs, survey schedule, and resources.

After the data is collected from the respondents the appropriate statistical test/method will be selected for data analysis.

Review of Literature

A. Association studies

Pearson correlation: The relationship between two continuous variables will be determined using Pearson correlation. The Pearson correlation also indicates the strength and direction of a linear relationship. The test coefficient *r* ranges from -1 indicating a perfect negative relationship to +1 for a perfect positive relationship with a value ‘0’(zero) with a relationship between two continuous variables (Cohen, 1988; Myeres et. al., 2010).

Point-Biserial correlation: The point-biserial correlation coefficient, rpb, will determine the strength of a linear relationship among one continuous variable and one nominal variable with two categories (Yes/No). The value can range from -1 to +1. The larger values indicate Stronger relationships, with zero (0) indicates no linear relationship among the two variables. The proportion of variance in one variable is explained by another variable (rpb2) (Chen and Popovich, 2002).

Spearman's rank-order correlation results in a coefficient, rs or ρ which measures the measure of the strength and direction of the association between two continuous or ordinal variables. For example, a researcher can determine the association between occupational stress and its effect on performance. Occupational stress can be measured on a continuous scale and performance effect can be measured on the ordinal scale— low effect, moderate effect, and high effect (Spearman, 1904).

B. Data/Dimension reduction

The Principal components analysis (PCA) is a variable/factor-reduction method and shares many similarities to exploratory factor analysis. PCA is used to reduce a larger set of variables into a smaller set of 'artificial' variables, the principal components that account for most of the variance in the original variables (Thurstone, 1947).

C. Predictions

Linear regression: The linear regression in simple terms measures the linear relationship between two continuous variables to predict the value of an outcome variable based on the value of an independent/predictor variable. Furthermore, linear regression measured whether the relationship is statistically significant; provides the amount of variation in the outcome/dependent variable is explained by the predictor/independent variable; (c) provides a direction and magnitude of any relationship, and (d) predict values of the dependent variables based on different values of the independent/predictor variable. A researcher can use linear regression to predict performance effect (dependent variable) based on the occupational stress an employee experiences (independent variable).

Multiple regression analysis: Multiple regression analysis predicts a continuous outcome variable based on multiple predictor/independent variables. This is an extension of simple linear regression. The multiple regression measured the overall variance, for and associated contribution of each of the predictors to the total variance explained. For example, one can measure an employee's performance using occupational stress factors like workload, role ambiguity, peer relation, and so on (Gelman and Hill, 2007).

Binomial logistic regression: A binomial logistic regression or logistic regression predicts the probability that an observation falls into one of two categories of a dichotomous dependent variable based on one or more independent variables that can be either continuous or categorical. The logistic regression is similar to linear regression, however, the probability of being in a particular category of the dependent variable, given the independent variable is predicted.

A researcher can binomial logistic regression to predict whether students will successful or unsuccessful in an exam based on the amount of tuition time, Maths is their basic/standard and exam stress. Here, your dichotomous dependent variable would be "exam performance", which has two categories – "successful" and "unsuccessful" – and you would have three independent variables: the continuous variable, "tuition time", measured in hours, the dichotomous independent variable, "Maths Basics", which has two categories – "yes" and "no" – and the ordinal independent variable, "exam stress levels", which has three levels: "low stress", medium stress" and "high stress" (Fox, 2016).

With this introduction, the following objectives are proposed

- To provide the basis for selecting appropriate variable—continuous, nominal, ordinal or ratio based on the objective
- To provide the basis for the selection of the appropriate statistical method based on the objective and data
- Interpret the results

D. Hypotheses

The authors assume that the researchers who carry out survey studies will provide appropriate hypotheses and in general null and alternate hypotheses are set out. However, But, any hypothesis is sufficient either null or alternative.

Generally, the following is an example of writing hypotheses

H_{11}: Effect of occupational stress and remote working is statistically significant significantly on the psychological well-being of an employee during Covid-19 Pandemic

H_{o1}: Effect of occupational stress and remote working is not statistically significant significantly on the psychological well-being of an employee during Covid-19 Pandemic

Or the following one will work
H_{12}: There are significant gender and age differences among the respondents on occupational stress and remote working factors effecting affecting the psychological well-being of the employees in the Information Technology sector

E. Theoretical Framework

The theoretical framework proposed, for this study on survey research was presented in Figure 1. Several statistical kinds of literature, papers, reports were considered for arriving at the framework and the framework was developed considering the literature and information provided by Dr Adam Lund (2020) @ Leard Statistics (Adam Land, 2020).

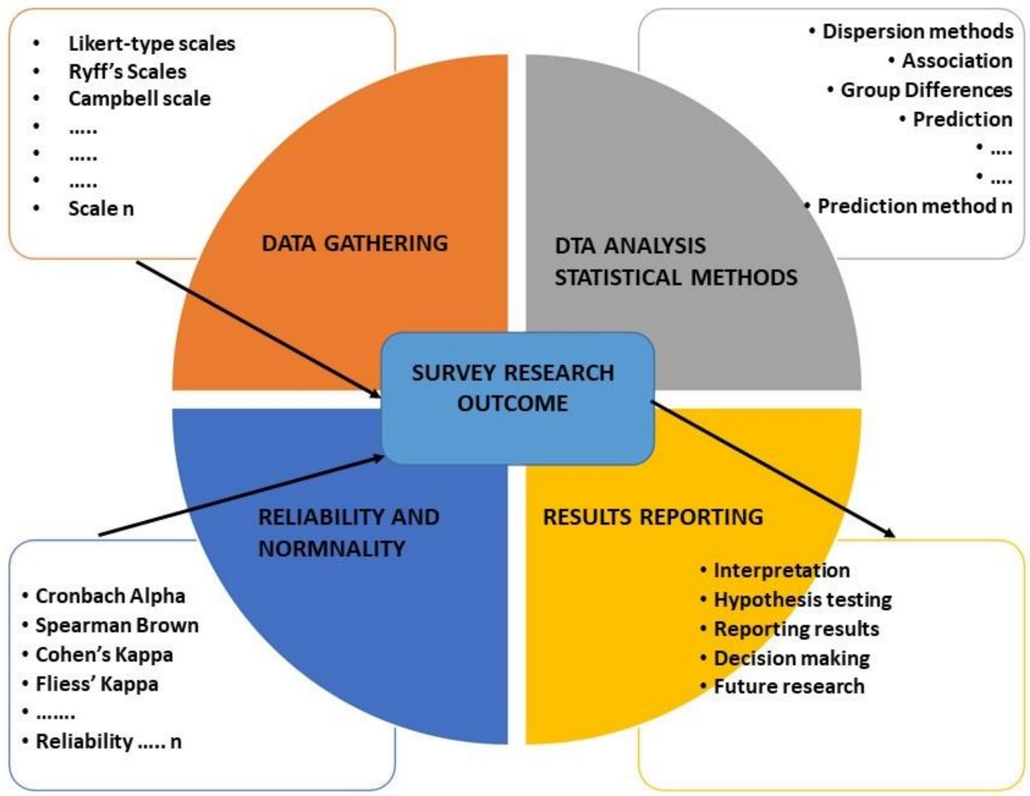

Figure 1. The theoretical framework of survey research

Data Analysis

F. Selection of appropriate variable

The respondent's data has four levels of measurement ordinal, interval, nominal, and ratio. It is a common practice that predictors like age, height should be measured on continuous variables, and some predictors like level of stress effect "low", "moderate" and "high" are measured using ordinal variables. However, several researchers knowingly or unknowingly treat ordinal variables as continuous as most such types of variables are independent or predictor variables. The Likert-type (Albaum, 1997)five-point scale has measurements which are: Strongly agree, Agree, Neutral, Disagree and Strongly Disagree to treat these variable as continues through the researchers does not know home many "times" of Agree is "Strongly Agree" and how many time of Disagree is Strongly Disagree. If we consider the ordinal items use categories that are not equally spaced Like "every day", "sometimes in a week", "twice a week", "once in the quarter", "never" and representing these variables as continuous is correct or incorrect? However, David Pasta (2009)emphasized how ordinal categories can be measured on a continuous scale, if a researcher makes a strong assumption that the successive categories of ordinal independent variables are equally spaced, For example, having no additional knowledge of another subject to a little knowledge than the teaching one (suppose if a teacher is teaching science, he does not know about teaching English, social sciences) will not have the same effect on the increase in student productivity to having full knowledge in another subject. Therefore, to use an ordinal variable as a continuous variable a researcher needs to test if any information is lost when considering the association between independent and dependent variables (Pasta 2009, William, 2016). Therefore, it is fine to use the ordinal variable as a continuous variable with minimal linear effects.

G. Selection of sample size

There are several methods of sample size estimation including ready-to-use tables, etc. however, most researchers (Cochran,.1977) for estimating the sample size from an unknown population.

Estimation of sample size for unknown population: the researchers used Cochran (1977) formula to estimate the sample size for this empirical study.

$$n_o = \frac{z^2pq}{e^2}$$

where no is the sample size, z is the selected critical value of desired confidence level, p is the estimated proportion of an attribute that is present in the population, q 1 p and e is the desired level of precision and this formula was used as IT sector where the population is unknown assuming the maximum variability which is equal to 50% (p=0.5) and taking 95% confidence level with ±5% precision, the required sample size is:

p = 0.5 and hence q = 1-0.5 = 0.5; e = 0.05 and z = 1.96

$$n_o = \frac{(1.96)^2(0.5)(0.5)}{(0.05)^2} = 384.16 = 384$$

The sample size for known population: For finite sample size i.e if the population is known. For example, in the case of Medical colleges in Hyderabad city where the respondents were approached the researcher used Yamane's formula calculating sample size:

As per Yamane's (1967) formula, a 95% confidence level and p = 0.5 size of the sample should be

$$N = \frac{N}{1+N(e^2)}$$

Where, N is the population size (i.e. total employees in Medical Colletes) and e is the level of precision. N = 700 with ±5% precision, assuming 95% confidence level and p = 0.5 the sample size is

$$n = \frac{700}{(1+700(0.05)^2)} = 254$$

H. Reliability of the survey instrument/questionnaire

The most commonly used reliability statistic is Cronbach's alpha and in some cases Spear Brown split-half reliability being used. However, it is up to the researcher to decide which reliability statistic will suit his/her study. Now the question is what should be the minimal value (Cronbach, 1951). Most researchers consider >0.6 Cronbach's alpha value will be reliable to go ahead for further analysis. However, it is not mandatory about the reliability of >06 to test hypotheses and the real issue in hypothesis testing of a relationship is validity. Reliability of your survey and measurement instrument, if it's independent or dependent variable will create an upper bound on validity, however. The relationship between validity and reliability can be expressed as:

rxy <= sqrt(rxx)

If you find the relationship, then you have validity and the reliability was sufficient. The reliability issue needs to be taken care when if a study fails to find the relationship. This does not mean low reliability means that there is no relationship of a failed relationship. Much more modern treatment is provided by Item Response Theory, specifically Rasch models. The Rasch model gets at the conceptual issue of reliability by examining the ability of the scale to separate persons and or items. Much more information is available through Rasch analyses than in classical test theory (http://www.winsteps.com). Englehard (2012) is good for comparing test score theory scaling theory applied to measurement. Boone et. al. (2014) narrated the construction and interpretation of graded response items such as Likert-type scaled items in building and testing invariant scales. A very accessible introduction to Rasch models in the dichotomous case and in the rating scale case and other multiple response graded items case were provided by Wright (1982, 1969) books.

Measurement of data: The researchers use appropriate scale survey data based on the study and the most common scale used is the Likert-type scale with 5 or 7 point item, and the measurement values vary from Strongly agree to Strongly agree. Ryff's (1995) psychological well-being scale is a 7-point scale and the measurement Strongly agree = 7, Somewhat agree = 6, A little agree = 5, Neither agree nor disagree = 4, A little disagree = 3, Somewhat disagree = 2, Strongly disagree =1. Several researchers use a 9 point scale ranging +4 to -4 to assess managerial

behavior, performance, and effectiveness. However, we need to convert all the scale son type of scale— a five-point, seven-point scale, and so on for ease of doing calculations.
Linear transformation proposed by IBM is the method of transforming different types of scale to a common scale (IBM Support, 2020).

Tables 1-5 present the different types of statistics that can be run on your data based on your type of variable and study design. The Association and relationship, prediction and relationship, group differences, reliability, and one sample test are presented respectively in table s1-5.

I. Linear regression - Interpretation

I am presenting here a brief note on how to interpret the liner regression coefficients both standardized and unstandardized on parameters Table 1.

Table 1. Regression coefficients for occupational stress, remote working and psychological wellbeing (n=400)[a]

Model	Factors	Unstandardized Coefficients	Std. Error	Standardized Coefficients	t	Sig.
		B		Beta		
1	(Constant)	1.598	0.226		7.076	0.000
	Remote Work	0.060	0.072	0.060	0.828	0.408
	Workload	0.039	0.056	0.038	0.699	0.485
	Peer77	0.177	0.048	0.202	3.675	0.000
	Physiological Factors	-0.057	0.068	-0.052	-0.845	0.399
	Role Ambiguity	-0.268	0.061	-0.266	-4.385	0.000
	Organizational Climate	0.393	0.083	0.382	4.726	0.000
	Psychological	-0.097	0.062	-0.111	-1.554	0.121
	Job Satisfaction	0.340	0.063	0.380	5.372	0.000

a. Dependent Variable: Psychological wellbeing

Interpretation: From the multiple regression analysis (Table 1) the independent occupational stress factors peer, role ambiguity, organizational climate, and job satisfaction are statistically significant and influencing the outcome variable of psychological wellbeing (Please refer to the Sig. column). The coefficient value of factor Peer 0.177 (**Value under Unstandardized Coefficeint**) represents the change in the dependent variable psychological wellbeing for one **unit** change in the independent variable overall. For one **unit** of increase of stress due to peer 0.177 **units** of psychological wellbeing will be affected. If we consider **standardized coefficients** a beta value of 0.202 indicates that a change of one **standard deviation** in the independent variable occupational stress results in a 0.202 standard deviations psychological wellbeing will be affected. **Standardized deviations are the units for Standardized beta coefficients.** Similarly, one unit increases stress due to organizational climate 0.393 units of psychological wellbeing will be affected, and considering the standardized beta value of 0.382 units, that a change one standard deviation in independent variable causing stress 0.382 standard deviation psychological wellbeing will be effect and son on. Therefore, psychological well-being can be predicted as:

Psychological wellbeing= $1.598+0.060_{(\text{remote work})}+0.039_{(\text{workload})}+0.177(\text{Peer})- 0.057_{(\text{physiological factors})}-0.268_{(\text{Role Ambiguity})}-0.393_{(\text{organization climate})}-0.097_{(\text{Psychological factors})}+0.340_{(\text{job satisfaction})}+$

J. Multinomial Logistic Regression Why?

Multinomial logistic regression is used to predict a nominal dependent/outcome variable with more than two categories given one or more independent variables.

Logistic regression is used to test the hypotheses about associations between a categorical dependent variable and one or more continuous or categorical independent variables. In a simple linear regression for one continuous independent variable *X* (an employee's occupational stress level) and one dichotomous outcome variable *Y* (the effect of employee performance) and the plot of such data results in two parallel lines, each corresponding to a value of the dichotomous outcome (Figure 1). Therefore, the two parallel lines are difficult to be inferences in OLS equation due to the dichotomy of outcomes. Therefore, the creation of categories for the predictor and compute the mean of the outcome variable for the respective categories will solve this problem. The resultant plot of categories' means will appear linear in the middle results S-Shaped curve (Figure 2, the S-shaped curve) and is very difficult to interpret as the extremes do not follow the linear trend, and the errors are neither normally distributed nor constant across the entire range of data (Peng, Manz, & Keck, 2001). Logistic regression solves these problems by applying the logit transformation to the dependent variable. In essence, the logistic model predicts the logit of *Y* from *X. Where is* and multinomial logistic regression is an extension of binary logistic regression

	Gender		
Remedial reading instruction	Boys	Girls	Total
Recommended (coded as 1)	73	15	88
Not recommended (coded as 0)	23	11	34
Total	96	26	122

Table: Sample data for Gender and Recommendation for remedial reading instruction

Source: Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The journal of educational research*, *96*(1), 3-14.

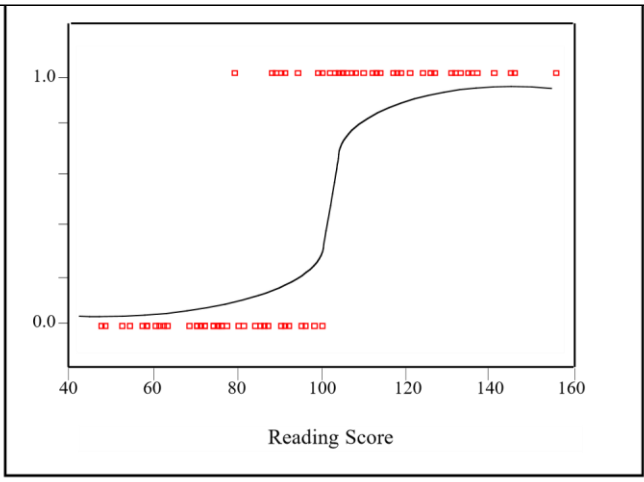

Figure 2. Relationship of a dichotomous outcome variable Y (1= Remedial Reading Recommended, 0= Remedial Reading Not Recommended) With a Continuous Predictor, Reading Scores

The multinomial logistic regression (MLR) to predict categorical variables with more than 2 outcomes which refer to polychotomous used to test multivariate associations. This regression is an extension of a chi-square analysis of three or more categorical outcome variables. In the MLR regression analysis, a reference category is selected from categorical outcomes of multilevel and successive logistic regression analysis conducted for each level of the outcome in comparison to the reference category. The resultant Odd ratios and 95% level confidence intervals are reported and referenced. Figure (3) represents the use of multinomial logistic regression. The predictor, gender, age, social support are being used to predict the outcome variable performance effect with more than 3 categories(low effect, moderate effect, and high effect)

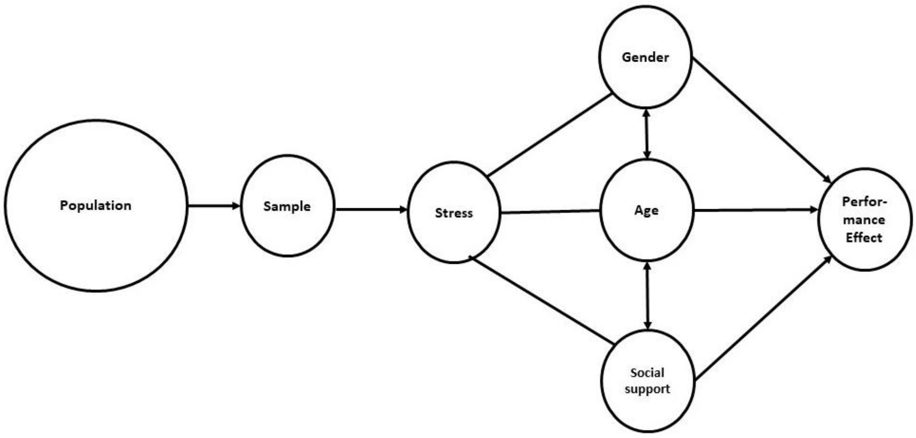

Figure 3. Multinomial logistic regression is used to predict for a polychotomous categorical outcome (Model based on Scale (2016) – www.scalelive.com

A brief notes on interpretation on multinomial logistic regression results. We are here reproducing the unpublished data of Prasad et al. 2020, 2021. The following is the parameter table which is an output of SPSS version 26. For easy interpretation, this table was converted into Table 2. Here, the High effect is the reference category and is compared with two other categories Low effect and the Moderate effect. In Table 1 Exp(ß) is nothing bur Odd Ratios (OR). Please note the following row in Table 2. From the table, if you see the Exp(ß) value of social support = 8.811 and from the significance column (.003), this 2.125 is lower bound and 36.541 is upper bound these values simple represented in Table 2 OR8.8(2.1-3.65)*** (***indicate *p*<0.001) and so on.

	LOW EFFECT		HIGH EFFECT	
Social support	OR 8.8(2.1-36.5)***	0.7	OR 2.0(1.14-3.75)***	0.30

Table 1: parameter estimates to predict Psychological wellbeing with organizational, supervisor, family and social supports (Source Prasad et al. 2020)									
								95% Confidence Interval for Exp(B)	
Psychological well-being Effect[a]		B	Std. Error	Wald	Df	Sig.	Exp(B)	Lower Bound	Upper Bound
Low Effect	Intercept	11.565	3.192	13.126	1	.000			
	Supervisor	.179	.615	.085	1	.771	1.196	.358	3.990
	Organization support	-1.165	.639	3.327	1	.068	.312	.089	1.091
	Social Support	2.176	.726	8.991	1	**.003**	**8.811**	**2.125**	**36.541**
	Family Support	-4.064	.806	25.437	1	.000	.017	.004	.083
	[Gender=1]	-.493	.666	.547	1	.459	.611	.166	2.254
	[Gender=2]	0[b]	.	.	0	.	.	.	.
	[Age_Group=1]	-2.529	1.059	5.703	1	.017	.080	.010	.635
	[Age_Group=2]	-22.698	.000	.	1	.	1.388E-10	1.388E-10	1.388E-10
	[Age_Group=3]	1.077	.945	1.299	1	.254	2.934	.461	18.690
	[Age_Group=4]	0b	.	.	0	.	.	.	.
	[Education=1]	-2.802	.761	13.549	1	.000	.061	.014	.270

	[Education=2]	0[b]	.	.	0	.	.	.	.
	[Discipline=1]	-.207	.000	.	1	.	.813	.813	.813
	[Discipline=2]	-.853	.754	1.279	1	.258	.426	.097	1.868
	[Discipline=3]	0[b]	.	.	0	.	.	.	.
	[Design.=1]	.064	..676	.009	1	.924	1.067	.284	4.009
	[Design.=2]	-20.707	.000	.	1	.	1.017E-9	1.017E-9	1.017E-9
	[Design.=3]	0[b]	.	.	0	.	.	.	.
Moderate Effect	Intercept	11.678	2.076	31.638	1	.000			
	Supervisor	.396	.380	1.085	1	.298	1.486	.705	3.129
	Organization support	-.994	.427	5.407	1	.020	.370	.160	.855
	Social Support	.727	.303	5.765	1	**.016**	**2.069**	**1.143**	**3.746**
	Family Support	-2.380	.478	24.764	1	.000	.093	.036	.236
	[Gender=1]	-1.297	.389	11.132	1	.001	.273	.128	.586
	[Gender=2]	0[b]	.	.	0	.	.	.	.
	[Age_Group=1]	-.811	.746	1.183	1	.277	.444	.103	1.917
	[Age_Group=2]	1.193	.682	3.060	1	.080	3.297	.866	12.553
	[Age_Group=3]	2.494	.717	12.083	1	.001	12.110	2.968	49.416
	[Age_Group=4]	0[b]	.	.	0	.	.	.	.
	[Education=1]	-1.961	.452	18.820	1	.000	.141	.058	.341
	[Education=2]	0[b]	.	.	0	.	..	.	.
	[Discipline=1]	20.219	5700.460	.000	1	.997	604134380.2	.000	.[c]
	[Discipline=2]	-1.524	.574	7.055	1	.008	.218	.071	.671
	[Design.=1]	.932	.481	3.756	1	.053	2.541	.990	6.524
	[Design.=2]	.597	.577	1.072	1	.300	1.817	.587	5.625
	[Design.=3]	0[b]	.	.	0	.	.	.	.

a. The reference category is: High Effect. B. This parameter is set to zero because it is redundant

Table 2. Presenting the results multinomial logistic regression: organizational, supervisor, family and social supports measured on dependent variable psychological wellbeing. Age, gender, education, discipline and designation results are also presented.

	Low effect(n=19)		Medium effect(n=159)	
Variable	OR(95% CI)	SE	OR(95% CI)	SE
Supervisor	1.20(0.4-4.0)	0.8	1.5(0.7-3.12)	0.4
Organization climate	0.30(0.09-1.091).	0.6	0.370(0.16- 0.855)**	0.4
Social support	8.8(2.1-36.5)***	0.7	2.0(1.14-3.75)***	0.30
Family support	.017(.004-.083)***	0.81	.093(.036-0.236)	0.48
Gender	.611(.166-2.254)	0.67	.273(0.128-0.586)	0.39
Age: 20-30	.080(.010-.635)***	1.1	.444(.1031.917)	0.75
Age: 31-40	Not Significant	0.0	3.297(.866-12.553)	0.68
Age: 41-50	2.934(0.461-18.69)	0.95	12.110(2.968-49.416)***	0.72
Education	.061(.014-.270)***	0.76	.141(.058-0.341)***	0.45
Discipline	Not Significant		Not significant	
Designation	Not Significant		Not Significant	

Note: Psychological wellbeing- Reference group: High effect(n=67) compared with low and moderate effect; OR=Odds Ratio, SE=Standard Error, 95% CI = Confidence interval, *p< 0.05; **p<0.01); *** p<0.001)

A Multinomial Logistic Regression was used to analyze the predictors on unordered group classification like low effect, moderate effect, and high effect in terms of psychological well-being. The reference category for the outcome variable was High Effect and the other two categories Low Effect and Moderate Effect were compared to this reference group. The main interest is of the current analysis was focused on the relational ship between social support, supervisory support, organizational support, and family support on psychological wellbeing (3 categories) while controlling age group, gender, discipline, and designation parameters (Table 2).

The first column of Table 2 is Low effect (referring to Psychological well-being Effect) was compared to the reference category High Effect (high psychological well-being effect). Referring to the parameter estimates of Table 12 the comparison will be done comparing low effect (first half). The results suggest that the predictor variable supervisor has no statistically significant influence on psychological wellbeing when compared with low effect keeping high effect as the reference category. Concerning organizational climate is significantly influencing the psychological wellbeing of an employee when compared with moderate effect (OR=0.370). For this model moderate effect versus high effect, for each unit increase in organizational support, the odds of increasing psychological wellbeing in moderate effect group is 0.370 times (95%, CI 0.16- 0.855) $p<0.01$); similarly, the predictor variable social support is statistically significant and influencing the outcome variable psychological wellbeing in both the groups (low and moderate) when compared with high effect group (Table 2).

The results from Table 12 indicate, for each unit increase in the social support odds of increasing psychological wellbeing of an employee in low effect group is OR 8.7 times (95%, CI 2.1 to 36.5), $p<0.001$), and for moderate effect, the group is OR 2.0 times (95% CI 1.14-3.75, $p<0.001$) and so on. Similarly family support in low effect group is OR 0.017(95% CI, 0.004-0.083, $p<0.001$); age group 20-3 OR 0.80 times (95% CI, 0.010-0.635, $p<0.001$) and age group 41-50 in moderate effect group OR 12.110 times (95%, CI 2.968-49.416, $p<0.001$)) when compared with high effect group. Similarly, the educational impact on psychological wellbeing is statistically significant for low effect OR 0.061 times (95%, CI 0.014-0.270, $p<0.01$), moderate effect OR 0.147 times (95%, CI 0.058-0.341, $p<0.01$). The variables discipline and designation no statistically significant effect on the outcome variable psychological wellbeing of Academician in higher education. Increased social support, family support, and organizational support will increase the psychological wellbeing of the Academician.

In general, where the odd ratios ex(ß) are <1.0 indicate moderate effects and >1.0 are significant and in this model social support, education, and age groups are strong predictors of psychological well-being of Academicians in higher education and Supervisor, discipline, the designation has no role influencing the psychological wellbeing.

Table 3. Different types of statistical tests/methods available for data analysis (Source Laerd Statistics)	
Association • Pearson's correlation • Point-biserial correlation • Pearson's partial correlation • Spearman's correlation • Kendall's tau-b • Goodman and Kruskal's gamma • Somers' d • Mantel-Haenszel test of trend • Cochran-Armitage test of trend • Chi-square test of association (2 x 2) • Chi-square test of independence (R x C) • Relative risk (2 x 2) • Odds ratio (2 x 2) • Goodman and Kruskal's lambda • Fisher's exact test (2 x 2 Independence) • Loglinear analysis **Survival analysis** • Kaplan-Meier **Data / dimension reduction** • Principal components analysis **Predictions** • Linear regression	**One sample** • One-sample t-test • Chi-square goodness-of-fit test **Reliability** • Cronbach's alpha • Cohen's kappa • Fleiss' kappa • Weighted kappa • Kendall's coefficient of concordance, W **MANOVA** • Hotelling's T2 • One-way MANOVA • Two-way MANOVA • One-way MANCOVA **Differences between groups** • Independent-samples t-test • Paired-samples t-test • One-way ANOVA • Two-way ANOVA • Three-way ANOVA • One-way repeated measures **ANOVA** • Two-way repeated measures **ANOVA** • Three-way repeated measures **ANOVA** • Two-way mixed ANOVA • Three-way mixed ANOVA (BBW)

• Standard multiple regression • Hierarchical multiple regression • Binomial logistic regression • Ordinal logistic regression • Dichotomous moderator, continuous IV	• Three-way mixed ANOVA (BWW) • One-way ANCOVA • Two-way ANCOVA • Mann-Whitney U test • Wilcoxon signed-rank test • Sign test • McNemar's test • Cochran's Q test • Kruskal-Wallis H test • Friedman test • Test of two proportions • Chi-square test of homogeneity (2 x C) • Chi-square test of homogeneity (R x 2) • Jonckheere-Terpstra test

Table 4. A list of appropriate statistical method or test for Association & Correlation

Study design	Number of variables	Type of variable	The statistical test
Association and Correlation	Two variables	Both Continuous	Pearson correlation Linear Regression
		Both ordinal	If you want to distinguish between an independent and dependent variable? Then the test you needed are: Somers' d and Jonckherre-Terpstra Test (ALTERNATIVE)
			If you DON'T want to distinguish between an independent and dependent variable? And Treat the ordinal variables as interval scaled? The test needed was Mantel-Haenszel test of trend
			If you DON'T want to distinguish between an independent and dependent variable? And don't want to Treat the ordinal variables as interval scaled? Then the test required was Spearman's correlation and Kendall's tau-b
		Both multinomial	Want to distinguish between an independent and dependent variable then Goodman and Kruskal's tau (τ) and Goodman and Kruskal's lambda (λ)
			Don't want to distinguish between an independent and dependent variable then: Chi-square test of independence and Fishers exact test
		Both Dichotomous	Want to distinguish between an independent and dependent variable then Test of proportions / Fisher's exact test, relative risk, odds ratio and fisher's exact test
			Don't want to distinguish between an independent and dependent variable then: Chi-square test of independence & Phi (φ) coefficient, Fisher's exact test, and Tetrachoric correlation (ALTERNATIVE)
		One dichotomous and one continuous	Point-biserial correlation Biserial correlation (Altenative)
		One dichotomous and one ordinal	Want to distinguish between an independent and dependent variable ordinal dependent variable carry out ordinal regression; if dichotomous dependent carry out Cochran-Armitage Test; and no distinction carry out Rank biserial correlation

		One continuous and one ordinal	Spearman's correlation and polyserial correlation
		One multinomial and one continuous	Eta (η) coefficient
		One multinomial an one ordinal	Row- or column-effects model
	Three+ variables with control variable	All continuous	Pearson partial correlation and multiple regression
		All ordinal	Ordinal logistic regression and partial gamma correlation
		All nominal both dichotomous main variables	Cochran-Mantel-Haenszel Test and Binomial logistic regression
		One or both multinomial	Multinomial Logistic Regression and Generalized Cochran-Mantel-Haenszel Test
		Continuous and ordinal	Pearson gamma correlation Pearson Partial correlation
		Continuous and dichotomous	Pearson partial correlation
	No control variable	all nominal	Loglinear analysis, binomial logistic regression, multinomial logistic regression
		Nominal and ordinal	Loglinear analysis, binomial logistic regression, multinomial logistic regression and ordinal logistic regression
		All ordinal	Ordinal logistic regression and log linear analysis

Table 5. A list of appropriate statistical method or test for Prediction and Relationships			
Study design	Type of dependent variable	How many independent variables	The statistical test
Prediction and Relationships	Continuous	One	Linear regression
		Two or more	Multiple regression
	Count		Poisson regression
	Ordinal		Ordinal logistic regression
	Dichotomous		Binomial logistic regression
	Multinomial		Multinomial logistic regression

Table 6. A list of appropriate statistical method or test – Group differences							
Study design	Type of study design	How many IV	How many groups your IV have	Type of DV	Have a covariate	Consider any DVs jointly	test
Group differences	Between subjects design	One	Two	Contin.	Yes	Yes	One-way MANCOVA
					Yes	No	One way ANCOVA
					No	Yes	Hotelling T^2
					No	No	Independent sample T test

			Two	Ordinal			Mann-Whitney U Test
				Dichotomous	Other variable yes		Binomial logistic regression
					No		Odds ratio, relative risk, Test of proportions, Fisher's exact test
				Multinomial			Chi-square test of homogenity
		Two			Yes	Yes	Two-way MANCOVA
					No	Yes	Two-way MANCOVA
					No	No	Two-way ANOVA
		Three			Yes	Yes	Three-way MANCOVA
					No	Yes	Three-way MANCOVA
					No	No	Three-way ANOVA
		Four or more			Yes	Yes	Factorial MANOVA
					No	Yes	Factorial MANOVA
					No	No	Factorial ANOVA
			Levels of Iv				
	Within subjects design	One	Two	Continuous		Yes	One-way Repeated Measures Manova
				Ordinal			Wilcoxon signed-rank test
				Dichotomous			McNamer's test
				Multinomial			Bhapkar's test Generalized estimating equations
			Three/ Four	continuous		Yes	One-way repeated measures MANOVA
				Ordinal			Friedman's test
				Dichotomous			Cochran's Q
				Multinomial			Generalized estimating equations
		Two		Continuous			Two-way repeated

							measures ANOVA
				Ordinal			Generalized estimating equations
				Dichotomous			Generalized estimating equations
				Multinomial			Generalized estimating equations
		Three		Continuous			Three-way repeated ANOVA
				Ordinal			Generalized estimating equations
				Dichotomous			Generalized estimating equations
				Multinomial			Generalized estimating equations
		Four of more		Continuous			Three-way repeated measures ANOVA
				Ordinal			Generalized estimating equations
				Dichotomous			Generalized estimating equations
				Multinomial			Generalized estimating equations
	Mixed Design	One between one within		Continuous	No	Yes	Three-way mixed MANOVA
					Yes	Yes	Three-Way Mixed MANCOVA

Table 7. A list of appropriate statistical method or test - Reliability			
Study design	Type of reliability	Variable type	The statistical test
Reliability	Internal consistency	Continuous	Cronbach'a alpha
		Ordinal	Ordinal Alpha
		Dichotomous	Kuder-Richardson (KR20) test
	Test comparison	Continuous	Intraclass correlation coefficient
		Ordinal	Weighted kappa (κ)
		Nominal	Cohen's kappa (κ)

Table 8. A list of appropriate statistical method or test – One sample test			
Study design	Type of variable	purpose	The statistical test

One sample	Continuous	Describing variable	Mean and standard Deviation
	Ordinal	"	Frequencies
	Dichotomous	"	Frequencies
	Multinomial	"	Frequencies
	Continuous	Making comparison	Kolmogorov-Smirnov one-sample test
	Ordinal	"	Chi square goodness of fit
	Dichotomous	"	Binomial test
	Multinomial		Chi square goodness of fit

Discussion

The selection and use of appropriate statistical tests depend on the survey and type of study variables. The researchers should report the results honestly from his/her analysis. For example, *r* value can range from -1 top +1 and if *r* value is negative and is nothing wrong, and it indicates a negative association. The researcher also chooses his/her study variables carefully based on the research study whether continuous, nominal or ordinal. Each statistical test has certain assumptions that need to be fulfilled and its researcher's responsibility to see that all the assumptions to carry out the test is met. For a correlation study, five assumptions need to be met and these are the basic requirements that should be fulfilled before running a statistical test. For ANOVA six assumptions need to be met— one independent variable should be measured on a continuous scale, the data should have independence of observations, no significant outliers in the data, the near-normal distribution of dependent variable for each group of the independent variable, and the data should be homogeneity of variances and so. There are standard procedures are available to test all the assumptions. In the case of multiple regression, eight such assumptions need to be met before running the test. To test one such assumption that data should independent of observations and this assumption is assessed by Durbin-Watson statistic and should have a value of >1. Another assumption is that residuals are normally distributed and this assumption is normally met by examining the Q-Q plots, and with a histogram superimposed normal curve and P-P plot.

Another important issue with ANOVA result is an ANOVA test can give the significance of the overall test, indicating there is differences in means but will not provide where those difference lie. Once the ANOVA is run and found significant results, a researcher can Tukey's HSD to find out which specific groups' means (compared with each other) are different for the significant groups. The SPSS system runs automatically Tukey-Kramer Method (Tukey, 1984) to take care of if you have unequal sample sizes. Rapid publication-ready MS-Word tables for one-way ANOVA will provide you the user creation of automatic post-hoc tables (Assaad, et al. 2014). A user can download free software at https://houssein-assaad.shinyapps.io/TwoWayANOVA/.

Acknowledgment

The authors sincerely thank Dr Adam Lund (2020) @ Leard Statistics. Every student will be immensely benefitted going through the website and information provided by the authors Dr Adam Lund at https://statistics.laerd.com/ which provides you how to run almost all statistical tests for survey research, all assumptions need to be met before running a test.

References

1. Adum Lund.(2020). https://statistics.laerd.com/spss-tutorials
2. Albaum, G. (1997). The Likert scale revisited. *Market Research Society. Journal.*, *39*(2), 1-21.
3. Andre Francis. 2008. Business Mathematics and Statistics. 6th Edition. South Western Cengage Learning EMEA, High Holborn House. 50-51 Bedford Row, London WC1R 4LR. ISBN 978-1-84480-128-
4. Annamalai, Sumathi., & Nandagopal, R. (2014). *Occupational Stress: A study of Employee Stress in Indian ITES Industry* (Vol. 1). Allied Publishers.
5. Assaad, H. I., Zhou, L., Carroll, R. J., & Wu, G. (2014). Rapid publication-ready MS-Word tables for one-way ANOVA. SpringerPlus, 3(1), 474.
6. Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Rasch analysis in the human sciences. Springer.
7. Chen, P. Y., & Popovich, P. M. (2002). Correlation: Parametric and nonparametric measures. Thousand Oaks, CA: Sage.
8. Cochran, W. G. (1977). Sampling Techniques: 3d Ed. New York: Wiley
9. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). New York: Psychology Press.

10. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, *16*(3), 297-334.
11. Engelhard, G. (2012). Invariant measurement: Using Rasch models in the social, behavioral, and health sciences. New York, N.Y. : Psychology Press
12. Fink, A. (2013). How to do sample surveys. The Survey Kit 2. 2nd Edition. Volume 1. ISBN-07619-2577-5. Sage publications. Thousand Oaks
13. Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. New York: Cambridge University Press.Fox, J. (2016). Applied regression analysis and generalized linear models (3rd ed.). Thousand Oaks, CA: Sage.
14. IBM support. Linear transformation to convert different scales to a common sale https://www.ibm.com/support/pages/transforming-different-likert-scales-common-scale
15. Myers, J. L., Well, A. D., & Lorch, R. F., Jr. (2010). Research design and statistical analysis (3rd ed.). New York: Routledge.
16. Pasta, D. J. (2009, March). Learning when to be discrete: continuous vs. categorical predictors. In *SAS Global Forum* (Vol. 248).
17. Peng, C. Y., Manz, B. D., & Keck, J. (2001). Modeling categorical variables by logistic regression. *American Journal of Health Behavior, 25*(3), 278–284
18. Prasad KDV and Mruthyanjaya Rao Mangipudi, Development of Occupational Stress, Coping, Motivation and Performance Scales: Data Analysis and Reporting, International Journal of Management, 11(8), 2020, pp. 1060-1074.
19. Prasad, K.D.V., Vaidya, R.W., Rao, M.M. (2020). Perceived psychological wellbeing of an academician in higher education is a function of organizational support, supervisor support, family support and social support during Covid-19 pandemic., *International Journal of Management*, 11(6), 2020, pp. 1292-1306.
20. Prasad, KDV, & Rao, M. (2021). A General Linear Model Approach: Development Of Psychological Well-Being, Remote Working, Employee Engagement, Job Satisfaction, Scales, Data Analysis And Reporting Concerning To Information Technology Sector. Journal of Contemporary Issues in Business and Government, 27(1), 1006-1035.
21. Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of personality and social psychology*, *69*(4), 719.
22. Scale. (2016). Multinomial logistic regression. http://www.scale.live/multinomial-logistic-regression.html
23. Spearman, C. (1904). The proof and measurement of association between two things. The American Journal of Psychology, 15(1), 72-101.
24. Thurstone, L. L. (1947). Multiple factor analysis. Chicago, IL: University of Chicago Press.
25. Tukey, J. W. (1984). *The collected works of John W. Tukey* (Vol. 1). Taylor & Francis.
26. Williams, R. (2016). Ordinal independent variables. *University of Notre Dame*.
27. Wright, B. D., & Masters, Geoff. (1982). Rating Scale Analysis. Chicago: Mesa Press.
28. Wright, B. D., & Stone, M.H. (1969). Best Test Design, Chicago: Mesa Press.
29. Yamane, T. (1967). Elementary sampling theory. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1967. Pp. x–405.