What Statistical Analysis Method Should I Use?

Table 1 shows common guidance for choosing a statistical analysis method based on the number and nature of dependent variables (sometimes referred to as outcome variables), the number and nature of your independent variables (sometimes referred to as predict variables) and the purpose of your research (1). Please note that these are general guidance and your data could be analyzed in multiple ways, each of which could yield reasonable answers. To apply the table correctly, you need to be able to identify the type of all the variables you are interested in the dataset (See figure 1). The statistical analysis methods can be done using R, SAS, STATA, and SPSS. If you any questions, please contact libdata@uwindsor.ca.

Table 1. General Guidance for Choosing a Statistical Analysis Method (2)

Number of Dependent Variables	Nature of Dependent Variables*	Number of Independent Variables	Nature of Independent Variables	Statistical Analysis Methods	Purposes
1	continuous & normal distributed	0	NA	One-sample t-test	Comparing one sample mean to a particular value
	quantitative & non-normal distributed			One-sample Wilcoxon signed rank test	Comparing one sample median to a particular value
	qualitative with 2 categories			Binomial test	Comparing a sample proportion to an expected proportion
	qualitative with 2 or more categories			Chi-square goodness-of-fit	Comparing the observed proportions to the expected proportions
	continuous & normal distributed	1	nominal with 2 levels (2 independent groups)	Two-sample independent t-test	Comparing two group means
	quantitative & non-normal distributed			Wilcoxon rank sum test	Testing if two samples are derived from the same population
	qualitative			Chi-square test	Testing if there is a relationship between two categorical variables
				Fisher's exact test	Testing if there is a relationship between two categorical variables when one or more of cells has an expected frequency of less than five

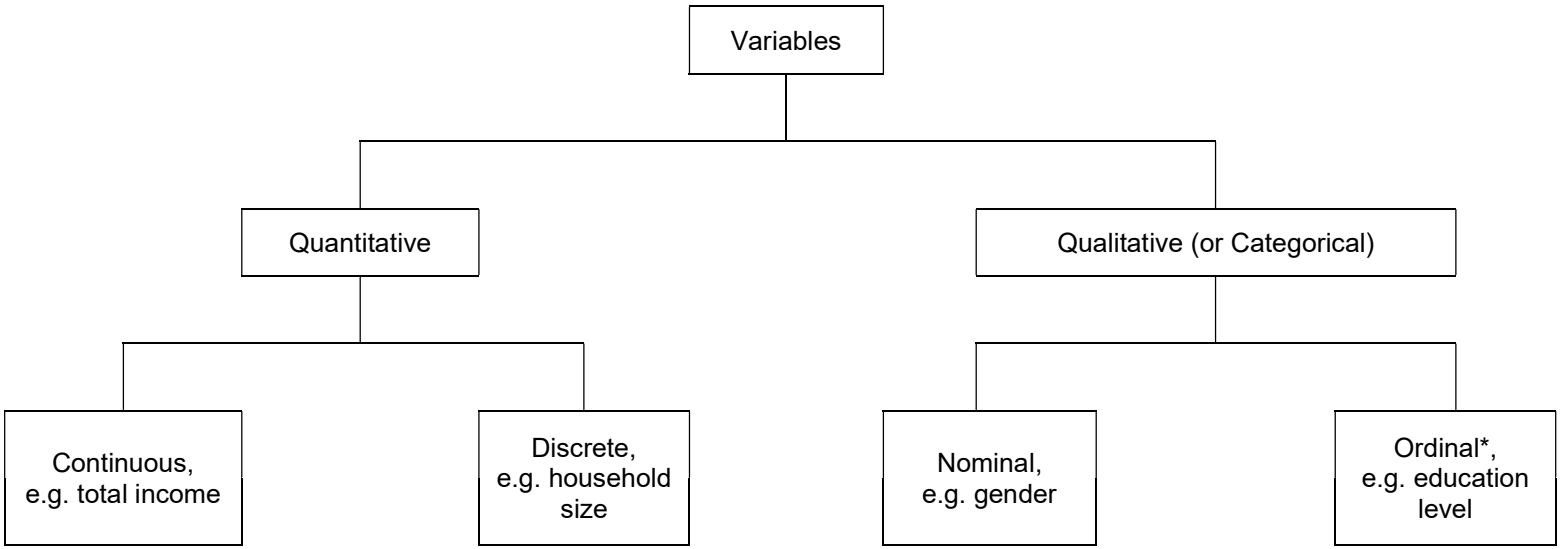
Number of Dependent Variables	Nature of Dependent Variables*	Number of Independent Variables	Nature of Independent Variables	Statistical Analysis Methods	Purposes
	continuous & normal distributed	1	nominal with 2 or more levels (independent groups)	One-way ANOVA	Comparing two or more than two group means
	quantitative & non-normal distributed			Kruskal Wallis test	Testing whether samples originate from the same distribution
	qualitative			Chi-square test	Testing if there is a relationship between two categorical variables
				Fisher's exact test	Testing if there is a relationship between two categorical variables when one or more of cells has an expected frequency of less than five
	continuous & normal distributed	1	nominal with 2 levels (dependent/matched groups)	Paired t-test	Comparing two means that are from the same subjects
	quantitative & non-normal distributed			Wilcoxon signed ranks test	Testing whether two dependent samples were selected from populations having the same distribution
	qualitative			McNemar's test	Testing if there are differences on a dichotomous dependent variable between two related groups
	continuous & normal distributed	1	nominal with 2 or more levels (dependent/matched groups)	One-way repeated measures ANOVA	Comparing two or more than two means that are from the same subjects
	quantitative & non-normal distributed			Friedman test	Non-parametric alternative to the one-way repeated measures ANOVA
	qualitative with 2 categories			Repeated measures logistic regression	Logistic regression that accounts for the effect of multiple measures from single subjects
	continuous & normal distributed	2 or more	nominal with 2 or more levels (independent groups)	Factorial ANOVA	Studying the effect of two or more independent categorical variables on the dependent variable
	quantitative & non-normal distributed			Generalized linear regression	Studying effect of two or more independent categorical variables on the dependent variable with non-normal distributed.

Number of Dependent Variables	Nature of Dependent Variables*	Number of Independent Variables	Nature of Independent Variables	Statistical Analysis Methods	Purposes
	qualitative with 2 categories			Factorial logistic regression	Explaining the relationship between dependent binary variable and two or more independent categorical variables
	qualitative with more than 2 categories			Multinomial or ordinal logistic regression	Explaining the relationship between one dependent nominal variable or ordinal variables and independent variables.
	continuous & normal distributed	1	continuous	Pearson correlation	Measuring the strength and the direction of a linear relationship between two quantitative variables
	continuous & normal distributed			Simple linear regression	Explaining the relationship between one dependent and one independent variable using a straight line
	quantitative & non-normal distributed			Non-parametric correlation	Measuring the strength and the direction of a linear relationship between two quantitative variables measured on at least an ordinal scale
	qualitative with 2 categories			Simple logistic regression	Explaining the relationship between one dependent binary variable and one independent variables
	continuous & normal distributed	2 or more	1 or more continuous variables and/or 1 or more qualitative variables	Multiple regression	Explaining the relationship between one dependent and two or more independent variables using a straight line
				Analysis of covariance	Comparing the mean of dependent variables in two or more groups considering variability of other variables, called covariates.
	qualitative with 2 categories			Multiple logistic regression	Explaining the relationship between dependent binary variable and more than one independent variables
				Discriminant analysis	Predicting group membership in the categorical dependent variable

Number of Dependent Variables	Nature of Dependent Variables*	Number of Independent Variables	Nature of Independent Variables	Statistical Analysis Methods	Purposes
2 or more	continuous & normal distributed	1	nominal with 2 or more levels (independent groups)	One-way MANOVA	Testing whether there are any differences between independent groups on two or more continuous dependent variable
		2 or more	any kinds of variables	Multivariate multiple linear regression	Explaining the relationship between multiple dependent variables and two or more independent variables
		0	NA	Factor analysis	Investigating whether a number of variables of interest are linearly related to a smaller number of unobservable factors
2 sets of 2 or more	continuous & normal distributed	0	NA	Canonical-correlation analysis (CCA)	Connecting two sets of variables by finding linear combinations of variables which have maximum correlation with each other

Note: *, Assumptions of normality concern model residuals, the deviation of the observed values from the model-predicted values. Many of these models are robust to violation of the assumption of normality, particularly in large samples.

Figure 1. Type of Variables (3)



Note: *, ordinal variables with five or more categories can often be treated as continuous (4).

References

1. McDonald JH. Handbook of biological statistics (3rd ed.). Baltimore, MD: Sparky House Publishing; 2014.
2. CHOOSING THE CORRECT STATISTICAL TEST. UCLA: Statistical Consulting Group [cited 22 July 2020]. Available from: https://stats.idre.ucla.edu/other/mult-pkg/whatstat/
3. Finlay B, Agresti A. Statistical methods for the social sciences. Dellen; 1986.
4. Sullivan GM, Artino Jr AR. Analyzing and interpreting data from Likert-type scales. *Journal of graduate medical education.* 2013 Dec;5(4):541-2.