f. Technical Jargons

g. Too many

h. Too long

i. Negative Qs. I do not like computer. If negative are used, then reverse the scoring at analysis.

2. Name the demographic variables: gender, age in years, age groups, education level, residence, experience, income etc.

3. Name the research variables: pain in flank, category of pain in throat, level of knowledge, level of attitude, level of practice, weight in Kg, height in cm, volume in ml, RBS in mg/dL, T3 level in pcg/ml 7. In case of categorical (nominal or ordinal) variable, tell the attributes (categories/ groups) of the variable a. Age grouping was; group 1 up to 50 years, group 2 more than 50 years in a study "prevalence of HTN in employees of a bank" b. Age in years was categorized as; group 1= 40-49, group 2 = 50-59, group 3 = 60-69, group 4 = 70 and above years for a study "prevalence of DM in adult age shopkeepers" c. The two attributes of residence were urban and rural. The five attributes of education level were: matric = group 1, graduation = group 2, masters = group 3, MPhil = group 4 and PhD = group 5e. Level of knowledge, level of attitude and level of practice (KAP) were determined by a questionnaire based on 5-point Likert Scale. There were so many questions for each of the three KAP variables with a range of 1-5 scores (strongly disagree, disagree, neutral, agree & strongly agree, respectively) for each question.

4. Identify independent, dependent, confounding and matching variables, where required9. Tell the data types (nominal/ordinal/interval/ratio); gender, residence and pain in flank were nominal data. Age groups, education level, and category of pain in throat were ordinal data. Age in years, level of knowledge, level of attitude, level of practice (all three on Likert Scale), pain score (on visual analogue pain scale-VAPS), weight in Kg, height in cm, volume in ml, RBS in mg/dL, T3 level in pcg/ml were interval/ ratio/ numeric/ continuous data.10. Attach Performa and questionnaire. if any.11. Mention which calculator or software was used for data analysis

5. **Data analysis plan**

Research is for a specified population; never for a sample. It is ideal to observe the entire population, but it is not feasible. Statistics as a discipline helps us in collecting data for a sample, analyze it for the sample (descriptive statistics; describe the sample) and then infer it on to the population from which it was drawn (inferential statistics; describe the population based on the

data collected from the sample). Inferential statistics includes estimation of parameter and hypothesis testing. Global literature is full of research articles which are restricted to sample, with no mention of the population. For us, it may be anything, but not research. Our authors have to give analysis plan for all the three components of the statistical analysis. It is widely stated and widely accepted narrative that the cross-sectional studies don't require hypothesis. It is a miss-understanding. Cross-sectional studies do require hypothesis. There may be some one dozen cross-sectional studies, each with many hypotheses published in this journal from 2018 to 2021, regarding burden/ magnitude (prevalence/ distribution) of malaria, leishmaniasis, DS-TB, DR-TB etc. Data analysis is simply a process of converting data (un-organized facts & figures) into information (organized facts & figures). Both qualitative and quantitative data are organized as per requirements of the topic and end users of the findings. When analyzed (organized), qualitative and quantitative facts and figures are mixed together to form a single piece of information or knowledge. There are two types of analysis.

Qualitative analysis

Qualitative data includes text, picture, audio and video. This analysis is based on qualitative argumentation (not included here).

Quantitative analysis

Quantitative data includes nominal, ordinal, interval and ratio data. This analysis is based on statistical computations (included here).

Descriptive analysis

It is the analysis of data collected from the sample. Here each variable is described separately without talking about its difference between the groups or within the groups or its relationships to any other variable in the same population.

Categorical (nominal and ordinal data) is analyzed by count and percentage. Numeric (interval and ratio) is subjected to tests of normality; Skewness, kurtosis, Kolmogoro-Smirnov test & histogram. If it is distributed normally; then it is analyzed by mean, minimum, maximum, range and SD. If it is distributed not normally (skewed); then it is analyzed by median (quartile 2), quartile 1 (Q1), quartile 3 (Q3) and Inter Quartile Range (IQR=Q3-Q1).

Inferential analysis: Here the data for the sample is inferred on to population. It includes estimation of parameters and testing of hypotheses.

Estimation of parameters

Here an interval in constructed around a sample statistic to estimate a parameter i.e. mean or proportion for a population at certain level of confidence, usually 95%. It is represented as