



APASL STC / PSSLD
Annual Meeting Islamabad
(Postgraduate Workshop)

Data Management and Analysis



By

Brig Aamir Ijaz (Retd)

MCPS, FCPS, FRCP (Edin) MCPS-HPEP DISEASES

Professor and Consultant in Chemical Pathology

Bahria International Hospital Rawalpindi


Ijaz_aamir@hotmail.com[[[

Objective

At the end of the present session participants will be able to describe and demonstrate data types and data handling respectively

PAKISTAN SOCIETY FOR THE STUDY OF LIVER DISEASES

Sequence of Presentation

- SPSS
 - Data Entry
 - Nature of Data
 - Descriptive Statistics
 - Tests of Significance
- 
- PSSLD
STUDY OF LIVER DISEASES



Statistical Package for the Social Sciences (SPSS)

- SPSS statistics (**Statistical Package for Social Sciences**) later modified to **Statistical Product and Service Solution** released its first version in 1968.
 - It is the most widely used program for *statistical analysis*.
 - It is used by *health researchers, education researchers, survey companies, marketing organisation* etc.
 - In 2009, SPSS is being owned by IBM and licensed version of IBM SPSS is made available.
-



Data Entry

RSSL D
RESEARCH CENTER FOR THE STUDY OF LIFE DISEASES

What is Variable?

A variable is a measurable characteristic of a person, object or phenomenon that can take on different values.

Examples:

- Age
- Gender
- Marital status
- Knowledge of Ethics
- Depression
- Lung Cancer



Types of Variables

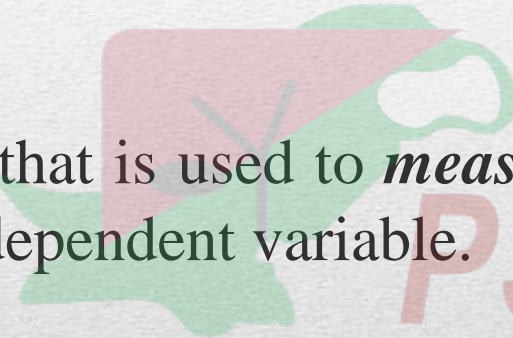
- Dependent Variable
- Independent Variable

Dependent Variable

- The variable that is used to *measure the problem under study* is called the dependent variable.
- A dependent variable is also known as a *Response variable* Or *Outcome variable*

PAKISTAN SOCIETY FOR THE STUDY OF LIVER DISEASES

PSSLD



Independent Variable

- The variables that are used to **explain the difference in the dependent variable** or to **cause changes in the dependent variables** are called the independent variables.
- An independent variable is also known as *Predictor variable* Or *Explanatory variable*

Task No 1

In a study of the association between Serum Triglycerides level and Non-Alcoholic Fatty Liver Disease (NAFLD). Which is dependent variable and which is independent variable?

Answer:

NAFLD: Dependent Variable (Yes / No data)

Triglycerides : Independent variable

Relative terms

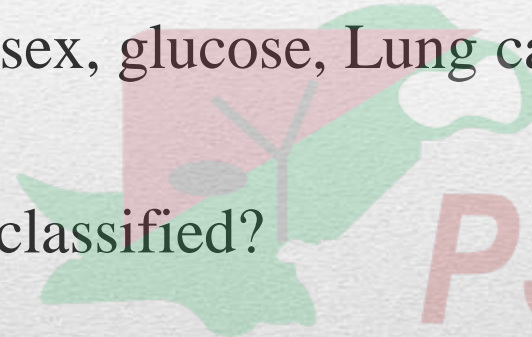
It depends on the Research Question that which variable is a Dependent Variable and which variables are an Independent ones.

Task No 2

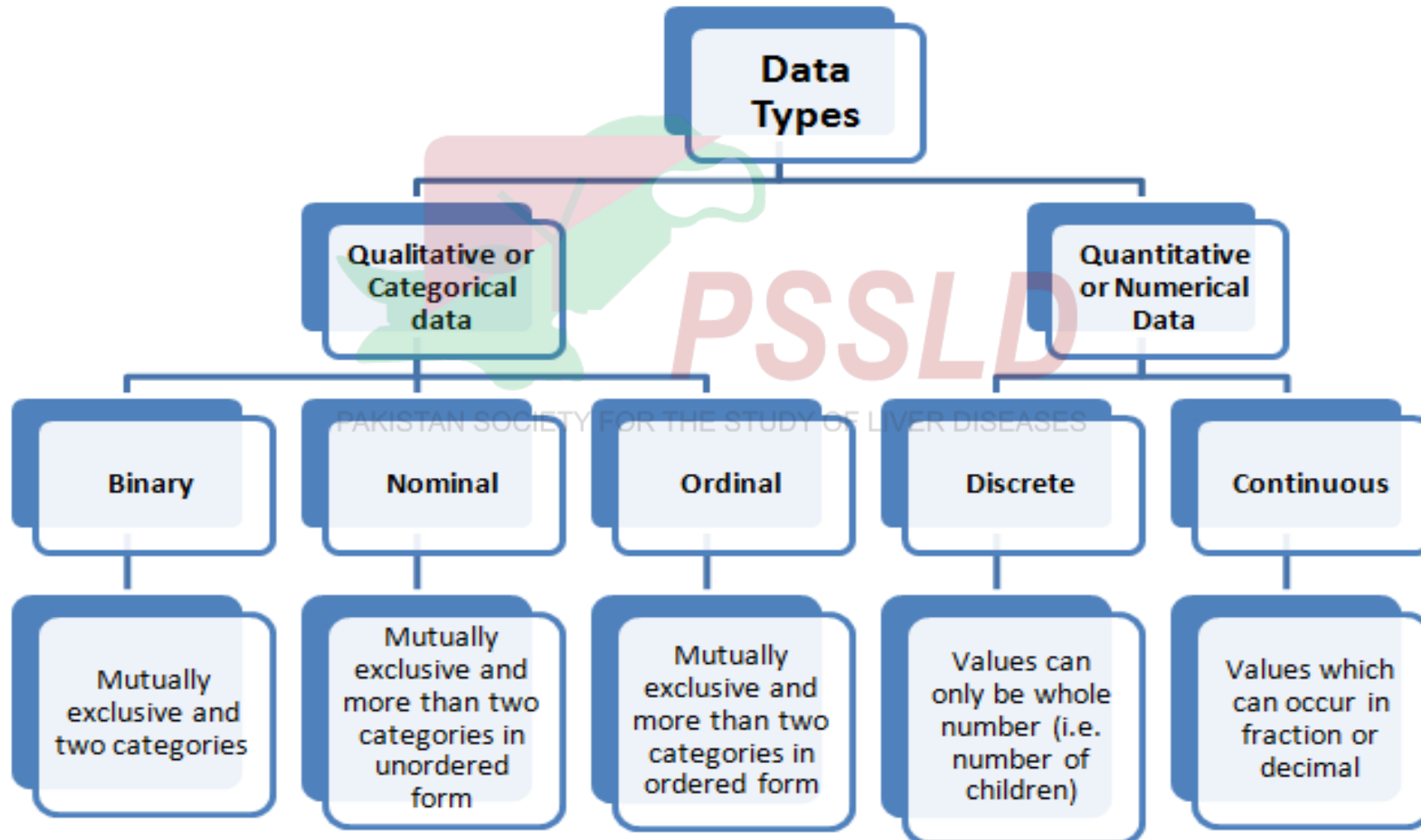
- Data are values of the observation recorded for variables (e.g. age, weight, sex, glucose, Lung cancer, blood pressure, etc.).
- How data is classified?
- **Qualitative or Categorical Data**
- **Quantitative or Numerical Data**

PAKISTAN SOCIETY FOR THE STUDY OF LIVER DISEASES

PSSLD



Classification of Data Types



Task 3

A frequency table showing the (hypothetical) distribution of Medical Students Professionalism Scores (expressed in *Grades 1-10*) of 100 students is as following:

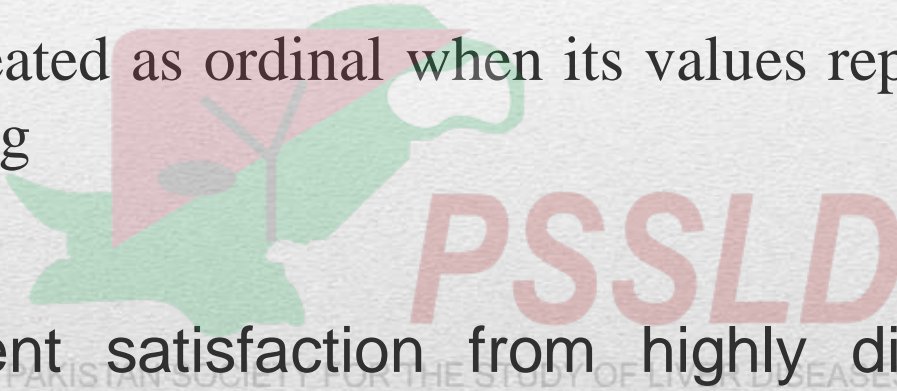
Result Grades	No of students obtaining the grade
3	10
4	11
5	19
6	25
7	28
8	12
9	4
10	1

Which type of variable it is?

Ordinal Variables

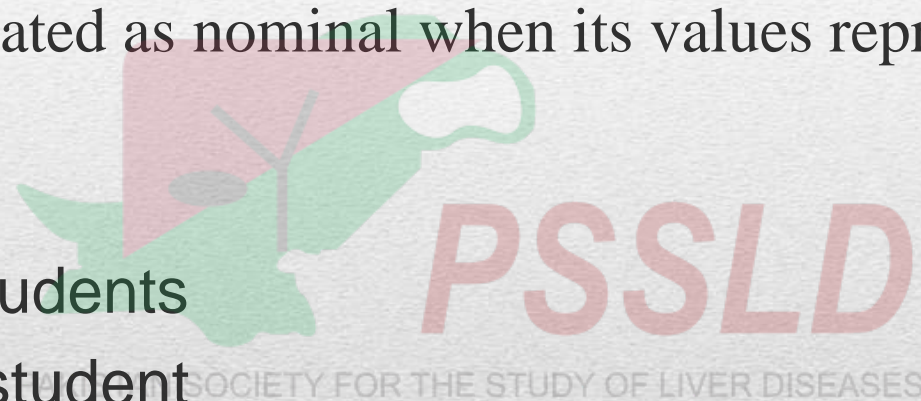
Ordinal Variable

- A variable can be treated as ordinal when its values represent categories with some intrinsic ranking
- **Examples**
 - ✓ Levels of student satisfaction from highly dissatisfied to highly satisfied.
 - ✓ Attitude scores representing degree of ethical practices



Nominal Variable

- A variable can be treated as nominal when its values represent categories with no intrinsic ranking
- **Examples:**
 - ✓ Gender of the students
 - ✓ Ethnic origin of student
 - ✓ Blood Groups
 - ✓ Roll No of students (Numeric)



Continuous Metric Variables

The metric variables represent data which is produced by proper measurement without any amount of guess.

Examples ??

- ✓ Blood Pressure
- ✓ Plasma Glucose
- ✓ Serum ALT
- ✓ Age of the Subjects



Discrete Metric Variables

- Whereas continuous metric data usually comes from *measuring*. *Discrete metric data, usually comes from counting.*

Examples ??

- ✓ Number of kids or number of siblings
- ✓ No of attempts allowed for FCPS Part II exam

PAKISTAN SOCIETY FOR THE STUDY OF LIVER DISEASES



Nature of Data

Nature of Data

- Symmetrical Data (Gaussian)
- Non-symmetrical



The First Step....

What should be the first statistical procedure to carry out ?

Tests of Normality?

- Shapiro Wilks Test: When $n = 3$ to 2000
- Kolmogorov-Smirnov: for those of $n > 2000$



Tests of Normality

What do these tests do?

- They compare the shape of sample distribution to the shape of a normal curve
- A significant test means the sample distribution is not symmetrical like a bell shaped curve

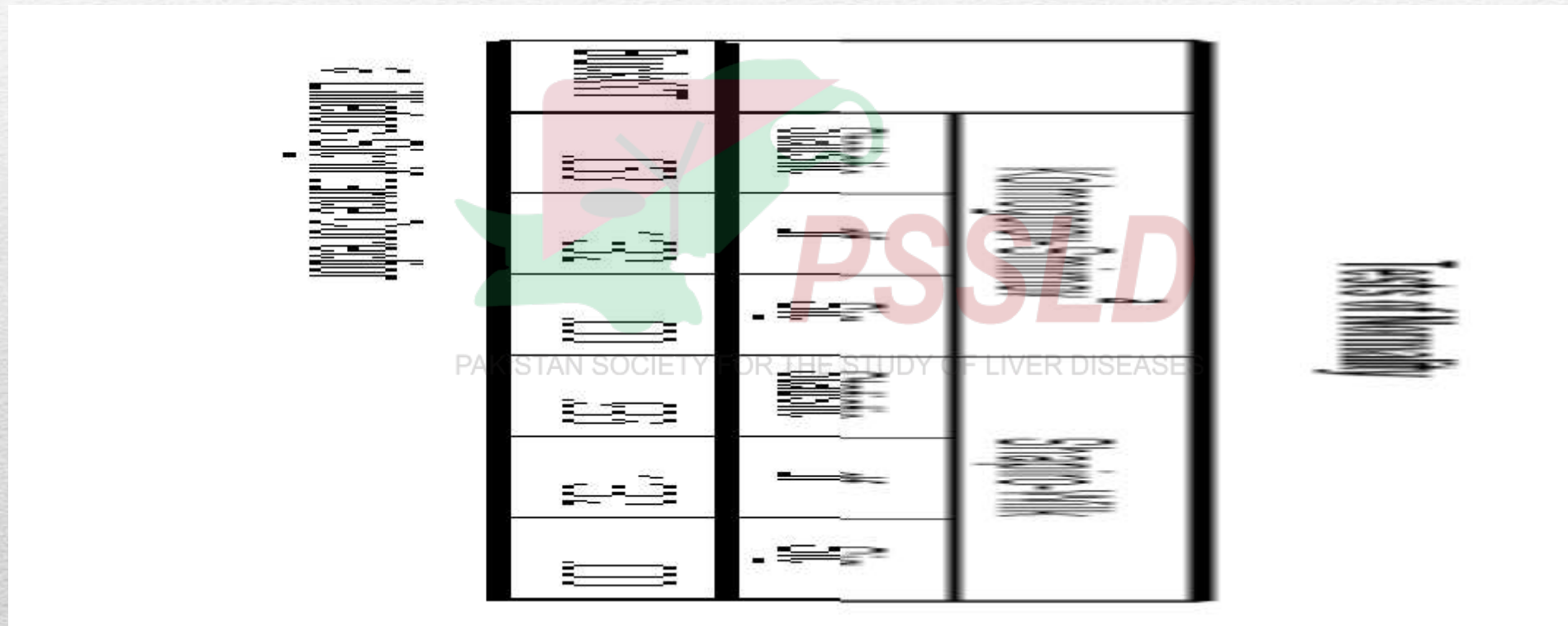


PAKISTAN SOCIETY FOR THE STUDY OF LIVER DISEASES



Task No 4

What is the nature of this data?



Non-Parametric (Asymmetrical)



Descriptive Statistics

PAKISTAN SOCIETY FOR THE STUDY OF LIVER DISEASES

Central Tendency

- Mean : For Symmetrical Data
(Mathematical average)
- Median : For non-parametric data
(Value at 50th Percentile after ranking the data)
- Mode: Most frequently occurring values



Measure of Scatter

- Standard Deviation: For Normal Data
- Interquartile Range (Single Value): For non-parametric data





Statistical Tests (Tests of Significance)

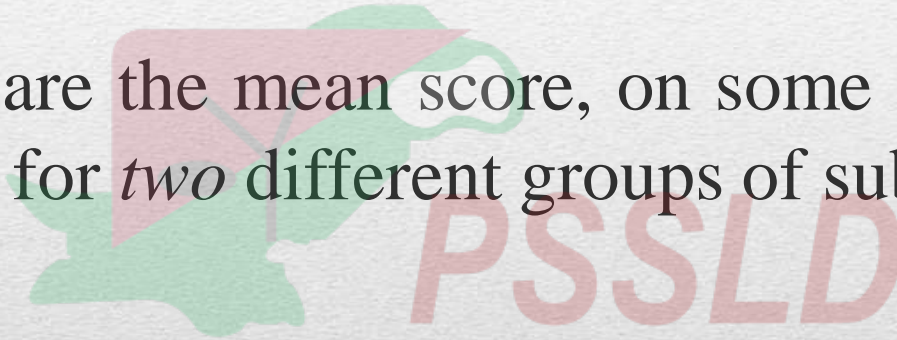
Tests for Parametric Data

- Independent Sample t Test (Student`s t Test)
- Paired Sample t test
- Single Sample t test
- One Way ANOVA



Independent Sampled t Test

- To compare the mean score, on some *continuous* variable, for *two* different groups of subjects.
- **Example:** Viral Load in two groups of patients on two different antiviral drugs



Paired Sampled t Test

- Used for continuous data of one group of people on two different occasions at Time 1 and Time 2.
- **Example:** Viral Load before and after one anti-viral drug in a group of patients

One Way ANOVA (Analysis of Variance)

- If we are interested in comparing the mean scores of more than two groups One Way ANOVA is the test of choice.
- **Example:** Viral Load in *three* groups of patients on *three* different antiviral drugs

One Way Repeated Measure ANOVA

- In a one-way repeated measures ANOVA design continuous variable is measured on the same continuous scale on three or more occasions.
- **Example:** Viral Load before, and after two anti-viral drugs in a group of patients used sequentially

Non-Parametric Tests

Parametric Test	Corresponding Non-parametric Tests
Independent Sample t Test (Student`s t Test)	Mann Whitney U Test
Paired Sample t test	Wilcoxon Matched Paired Test
One way ANOVA	Kruskal Wallis Test
One-way repeated measures ANOVA	Friedman Test

References

1. Bowers D. Medical Statistics from Scratch An Introduction for Health Professionals 2nd Edition. John Wiley & Sons Ltd, West Sussex, England. 2008.
 2. Harris M, Taylor G. Medical Statistics Made Easy. First Edition. Martin Dunitz, London, UK. 2003
 3. SPSS Log (online). Cited on 11 June 2011. Available from <http://www.spsslog.com/2006/05/03/nominal-ordinal-and-scale/>
 4. Likert scale. From Wikipedia, the free encyclopedia. (online). Cited on 11 June 2011. Available from http://en.wikipedia.org/wiki/Likert_scale
-

THANK YOU



Q & A

PAKISTAN SOCIETY FOR THE STUDY OF LIVER DISEASES
