

COMPUTER ASSISTED STATISTICS

Dr. Sujith J Chandy
Clinical Pharmacology Unit
CMC Vellore

INTRODUCTION

- **Various phases for a study :**

1. Identify a problem
2. Ask yourself a question
3. Define the aim with specific objectives
4. Choose a study design
5. Construct the methodology
6. Estimate the sample size
7. Data Collection
8. Analysis
9. Interpretation
10. Conclusion

- **Planning for a study:**

- You remember - Aim, design, methodology, sample size
- You forget - data management and analyses

- **Data management :**

- Data collection, entry, arrangement, processing

- **Analysis:**

- choice of analysis, process of analysis

- **Assistance from a friend:**

- Biostatistician, Colleague, or Software

Aim of talk :

Get to know statistical software - A process of sensitization

Areas of talk :

1. Data Management software
2. General software packages for analysis
3. Special function software packages

DATA MANAGEMENT

- **Spreadsheet :-**

- **Functions :**

1. Arithmetical - 4 signs, square root etc
2. Statistical - Averages, ranges, MMM, SDs etc
3. Data sorting - Arranging, distribution analysis
4. Data storing - 9999 rows and 255 columns - over 2 million cells
5. Transfer into other programs - word processors, database

Examples: Quattro Pro, Lotus 1-2-3,

Microsoft Excel -PopTools, XL statistics

STATISTICAL ANALYSIS

- Data analysis in multiple ways:
 - each of which could yield legitimate answers
- Choice of analysis method:
 1. the number of dependent variables (outcome variables)
 2. the nature of independent variables (predictors)
 3. the nature of dependent variable
 - whether interval variable, ordinal or categorical variable,
 - whether it is normally distributed

STATISTICAL PROGRAMS

- The Past - programs expensive, unfriendly
- The Present - budget driven (many programs available)
 - user friendly - menu driven, Windows/Mac graphic user interfaces
- Broad types of programs:
 1. General Packages: support a wide variety of statistical analyses
 2. Special function Packages: with a specific area of analysis
 3. Curve fitting and Modelling: handle complex, nonlinear models
 4. Surveys, Testing and Measurement: useful in business
and social sciences

GENERAL STATISTICAL SOFTWARE PACKAGES

Arcus Pro-Stat - useful statistical functions, user friendly

EpiInfo - downloadable, free, ease to use

Kwikstat - popular in USA

Minitab -available for DOS, Windows, Macintosh, Unix

StatSoft - Comprehensive statistical program

Statgraphics - Very powerful with integrated graphics

SYSTAT - General statistic package

UNISTAT - range of statistical analysis with charting capabilities

SAS

SPSS

STATA

EPI INFO

Features:

- Can rapidly develop a questionnaire or form
- Can customize the data entry process
- Can enter and analyze data
- Epidemiologic statistics, tables, graphs, and maps produced
- Simple commands - READ, FREQ, LIST, TABLES, GRAPH
- Epi Map displays geographic maps with data from Epi Info

STATA

**Statistical Software for
Professionals**

1984

General Features :

- both easy to learn and yet very powerful:
 - beginners and veterans
 - interactive style of operation
- commands give immediate results
- one line commands which can be entered:
 - one command at a time (a mode favored by beginners)
 - many at a time (a mode favored by power users)
- if you make mistake in Stata command:
 - easy to diagnose and correct the error

Data management :

- Data management capabilities not as extensive as SAS
- Powerful yet very simple data management commands
 - allows complex manipulations of data with ease
- Works primarily with one data file at a time
 - tasks that involve working with multiple files at once - difficult

Statistical Analyses :

- Performs most general statistical analyses :
 - regression, logistic regression, survival analysis
 - analysis of variance, factor analysis
 - computing risk, rate and odds ratios with confidence intervals
- Greatest strength - regression, logistic regression
- Excels - survey data analysis
- Greatest weakness -
 - analysis of variance
 - traditional multivariate methods (manova, discriminant analysis)

Graphics :

- Like SPSS, Stata graphics created using Stata commands or using a point and click interface
- Unlike SPSS, the graphs cannot be edited using a graph editor
- The syntax of the graph commands :
 - easiest of the three packages
 - the most powerful
- Stata graphs are high quality, useful in publications

Summary:

- Good combination of ease of use and power.
- Easy to learn
- Powerful tools for data management
- Cutting edge statistical procedures
- Easily download programs developed by other users
- Create own Stata programs that become part of Stata
- Compact in terms of disk space, computationally efficient

SAS

Software and Services

1976

General Features :

- Large general purpose package
- Power and programmability - Professionals love it
- Difficult to learn
 - (STATA more interactive, SPSS more trial based)
- To use SAS:
 - write SAS programs that manipulate your data and perform analyses
- If you make a mistake in a SAS program :
 - hard to see where the error occurred or how to correct it.

Data management :

- Powerful in data management
 - allows you to manipulate data
- Long time to learn and understand data management
- Complex data management tasks done
(simpler commands in Stata or SPSS)
- SAS can work with many data files at once
 - easing tasks that involve working with multiple files
- SAS can handle enormous data files - up to 32,768 variables

Statistical Analyses :

- Performs most general statistical analyses :
 - regression, logistic regression, survival analysis
 - analysis of variance, factor analysis, multivariate analysis
- Greatest strengths of :
 - ANOVA, mixed model analysis and multivariate analysis
- Greatest weakness:
 - ordinal and multinomial logistic regression
- analysis of survey data is quite limited as compared to Stata

Graphics :

- Most powerful graphic tools among all of the packages
- Technical and tricky to learn
- Graphs are created mainly using syntax language
- SAS 8 does have a point and click interface for creating graphs
 - not as easy to use as SPSS

Summary :

- SAS is a package geared towards professionals
- Steep learning curve and can be frustrating at first
- Powerful data management
- Range of statistical procedures wide
- Ability to work with numerous data files at once

SPSS

**Statistical Packages for Social
Sciences**

1968

General use :

- Beginners enjoy because
 - it is easy to use
- "point and click" interface
 - allows you to use pulldown menus to select commands
- SPSS does have a "syntax" language
 - syntax generally overly complicated and often unintuitive
- Popular in the social sciences and psychology

SPSS

File Edit Data View Windows Help

Statistics Display Diagnostics Window Help

Summaries
 Compare Means
 ANOVA Models
 Correlate
 Regression
 Classify
 Data Reduction
 Scale

Nonparametric Tests
 Multiple Comparisons

Chi-Square
 Homogeneity
 Linear
 Logistic
 1-Sample K-S...
 2 Independent Samples...
 K Independent Samples...
 2 Related Samples...
 K Related Samples...

	#	Label	Value
1	1	Market price	Over 1 year
2	2	Reliability of	10-12 months
3	3	Reliability of	7-9 months
4	4	Mail order	6-8 months
5	5	Mail order	5-6 months

SPSS: Procedures Manual

Data Management :

- Friendly data editor that resembles Excel
 - allows you to enter your data
 - and attributes of your data (missing values, value labels)
- SPSS does not have very strong data management tools
- SPSS primarily edits one data file at a time
- not very strong for tasks that involve working with multiple data files at once
- SPSS data files can have 4096 variables

Statistical Analyses :

- Performs most general statistical analyses:
 - regression, logistic regression, survival analysis
 - analysis of variance, factor analysis, and multivariate analysis
- Greatest strengths :
 - analysis of variance and multivariate analysis
- Greatest weakness :
 - absence of robust regressions
 - absence of survey data analysis

Graphics :

- Simple point and click interface for creating graphs
- once you create graphs
 - can be extensively customized via its point and click interface
- Graphs are very high quality
 - can be pasted into other documents (Word or Powerpoint)
- SPSS does have a syntax language for creating graphs
- Many of the features in the point and click interface
 - are not available via the syntax language
- Syntax language more complicated than that provided by Stata, but simpler than the SAS

Summary :

- Large and cumbersome, but easy to use
- As a professional - you may outgrow it over time
- Strong in the area of graphics
- Weak in cutting edge statistical procedures
- Hard to keep track of exactly what you've done

QUALITATIVE DATA

- Analytical process begins during data collection
- Data preserved in textual form and "indexed" to generate or develop analytical categories and theoretical explanations
- Software designed to enable complex organisation, retrieval of data
- Widely used - QSR NUD*IST and ATLAS.ti.
- EZ-TEXT-- a software program from C.D.C.
-to create, manage, analyze semi-structured qualitative databases

SPECIAL FUNCTION PACKAGES

- **Sample Size & Power :**

1. STPLAN, NQUERY, PEST, POWER, SAMPSIZE
2. SAMPLEXS - for crosssectional surveys
3. GPOWER - precision analysis for t-tests, Chi-square tests
 - computes power, sample sizes, alpha, beta, and alpha/beta ratios

- **Randomization :**

1. RANDOMIZATION GENERATOR - Windows application
 - simple, blocked randomisation, and gives output in spreadsheet
2. RANDOMIZATION.COM - on-line randomization program
3. STATSDIRECT - randomization into groups, matched pairs

- **Metanalysis :**

1. META-ANALYSIS 5.3 -- DOS statistics software
 - selects the analysis of exact p values or effect sizes
 - plots a display of correlation coefficients
2. EasyMA - Clinical Trial Results
3. EPIMETA - DOS based, but has Windows like interface

- **Person Years and Mortality Ratios :**

- PANCOMP

- **Diagnostic Efficiency Statistics :**

1. DIAGNOSTIC STATISTICS - - sensitivity, specificity, positive and negative predictive power, Kappa

- **Risk Estimation :**

1. ABSRISK - MSDOS

- for estimating absolute risks from relative risks
- Uses age-specific mortality and morbidity data to convert relative risk estimates into absolute risk estimates

- **Missing Data :**

1. MANET

- Macintosh software for interactive graphics tools for data sets with missing values.

- **Clinical Trial :**

1. CLINTRIAL 2. MACRO

- **Pharmacokinetic and Pharmacodynamic Data :**

1. BOOMER - Non-linear Regression Program

- Includes normal fitting, Bayesian estimation, with integrated or differential equation models

2. NONMEM - Population Kinetics software

- **Surveys :**

1. CSPRO (Census and Survey Processing System)

2. STATPAC Survey Software

- **Seasonal Adjustments :**

- ARIMA, DEMETRA

CONCLUSION

- Each package - its own unique strengths and weaknesses
- SAS, Stata and SPSS - wide variety of statistical analyses
- Stat/Transfer - converts data files from one package to another
- Switching from one analysis package to another :
 - depends on nature of problem
 - SAS: analyses using mixed models
 - Stata : logistic regression
 - SPSS: analysis of variance
- Choice of program :
 1. budget
 2. experience
 3. size of data
 4. complexity of analyses

- **Take home message for CaS:**

The right analysis for the right problem,
Your seers will have good views.

The right software for the right analysis,
Your peers will give good reviews!

Whichever software you choose, let there be no excuse,
Persist with its use, be familiar with the cues,
You're sure to get your dues,
And who knows, one day, you might just be... in the news!