SPSS for Quantitative Data Analysis

Conducted by

Dr. Prodhan Mahbub Ibna Seraj

Ph.D. (TESL)

Associate Professor

American International University Bangladesh (BELTA)

Email: mahbub.seraj@aiub.edu

WhatsApp: +8801727191342

Learning Outcomes

- Concepts of Research
- Conceptualizing Research Topics
- Searching Research Papers (Literature)
- Research Paper: Definition and Types
- Elements of a Research Paper
- Research Methodology
- Referencing (Mendeley)
- Finding Scopus Indexed Journals

What is SPSS?

- Software **Package used for Statistical Analysis** of data.
- Produced by SPSS Inc. **in 1968**.
- SPSS used to stand for "**Statistical Package for the Social Sciences**"
- Later changed to "**Statistical Product and Service Solutions**"
- Acquired by IBM in 2009. Now known as **IBM-SPSS Statistics**

- Current version is 22.0
- SPSS is a **commercial** software
- Statistic 17.0 (basic package) is freely available for WPI students
- Several specialized packages can be bought:
  - SPSS **Data Collection** (for surveys)
  - SPSS **Modeler** (for data mining)
  - SPSS **Analytic Catalyst** (for Big data) etc.

Supported File formats

- Basic format is .SAV
- Supports other common formats such as .XLSX, .CSV, .DAT etc
- SPSS syntax file (.SPS) can be used to convert other formats to SPSS format

Why use it then?

- Very rich collection of Statistical tests and methods
- Outputs an extensive set of metrics and statistically important factors
- Support available
- Well known in non-CS fields

- Data View
  - Used to display data
  - Columns represent variables
  - Rows represent individual units or groups of units that share common values of variables
- Variable View
  - Used to display information on variables in dataset
  - TYPE: Allows for various styles of displaying
  - LABEL: Allows for longer description of variable name
  - VALUES: Allows for longer description of variable levels
  - MEASURE: Allows choice of measurement scale
- Output View
  - Displays Results of analyses/graphs

Data Entry Tips

- For large datasets, use a spreadsheet such as EXCEL which is more flexible for data entry, and import the file into SPSS
- Give descriptive LABEL to variable names in the VARIABLE VIEW
- Keep in mind that Columns are Variables, you don’t want multiple columns with the same variable

Importing data into SPSS

To import an EXCEL file, click on:

FILE → OPEN → DATA then change FILES OF TYPE to EXCEL (.xls)

To import a TEXT or DATA file, click on:

FILE → OPEN → DATA then change FILES OF TYPE to TEXT (.txt) or DATA (.dat)

You will be prompted through a series of dialog boxes to import dataset

Reliability Test

- After Importing your dataset and providing names to variables, click on:
- ANALYZE → SCALE→ RELIABILITY ANALYSIS
- Choose the items to be analyzed and place them in the box on the right
- Click on OK
- Results window OPEN

Descriptive Statistics-Numeric Data

- After Importing your dataset, and providing names to variables, click on:
- ANALYZE → DESCRIPTIVE STATISTICS→ DESCRIPTIVES
- Choose any variables to be analyzed and place them in the box on right

Descriptive Statistics-General Data

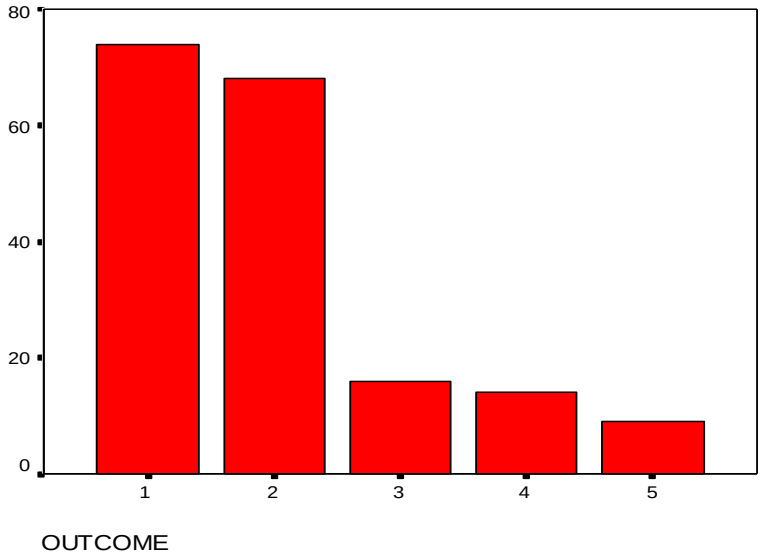
- After Importing your dataset, and providing names to variables, click on:
- ANALYZE → DESCRIPTIVE STATISTICS→ FREQUENCIES
- Choose any variables to be analyzed and place them in box on right
- Options include (For Categorical Variables):
  - Frequency Tables
  - Pie Charts, Bar Charts
- Options include (For Numeric Variables)
  - Frequency Tables (Useful for discrete data)
  - Measures of Central Tendency, Dispersion, Percentiles
  - Pie Charts, Histograms

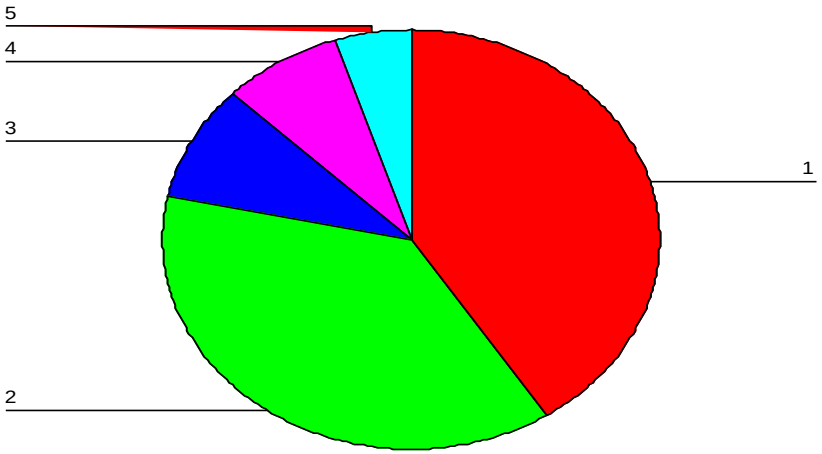
Vertical Bar Charts and Pie Charts

- After Importing your dataset, and providing names to variables, click on:
- GRAPHS → BAR… → SIMPLE (Summaries for Groups of Cases) → DEFINE
- Bars Represent N of Cases (or % of Cases)
- Put the variable of interest as the CATEGORY AXIS

- GRAPHS → PIE… (Summaries for Groups of Cases) → DEFINE
- Slices Represent N of Cases (or % of Cases)
- Put the variable of interest as the DEFINE SLICES BY

Example 1.5



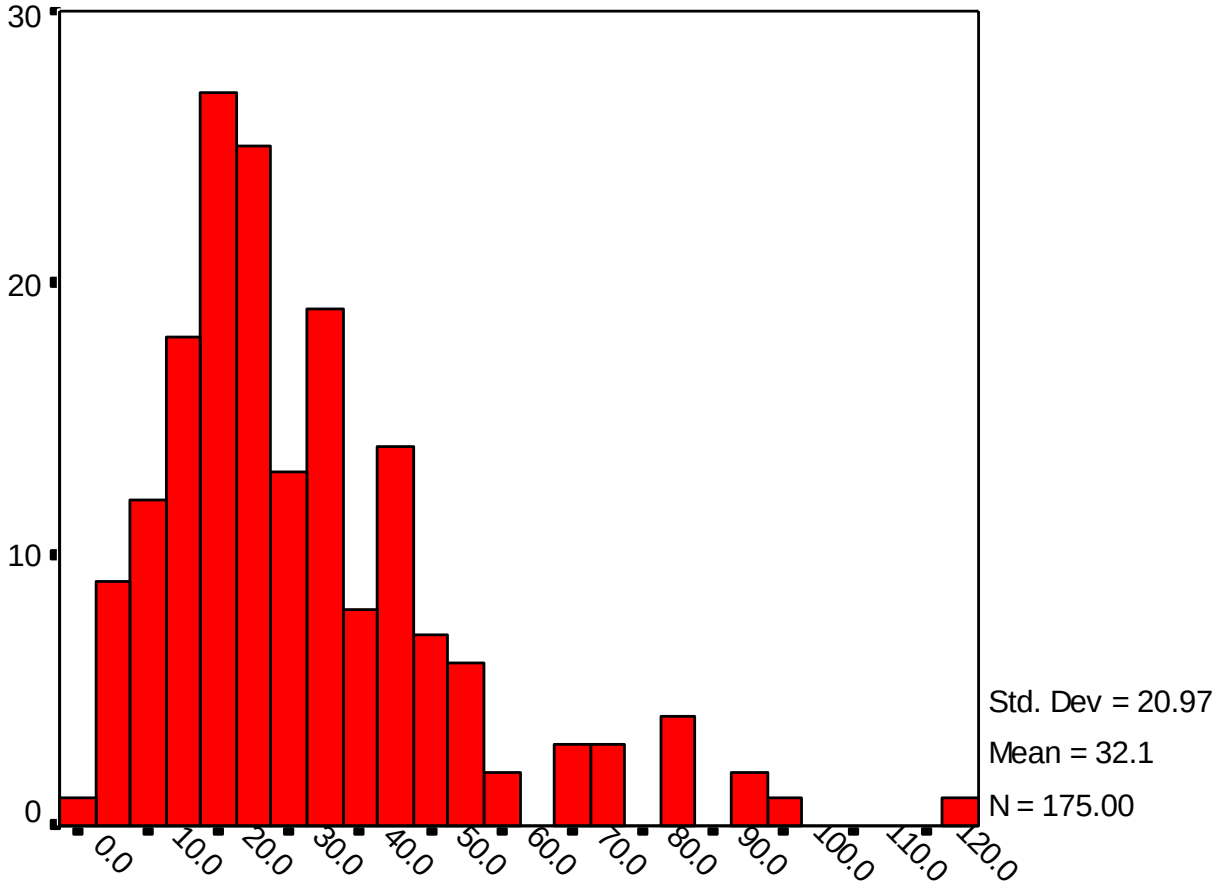


Histograms

- After Importing your dataset, and providing names to variables, click on:
- GRAPHS $\rightarrow$ HISTOGRAM
- Select Variable to be plotted
- Click on DISPLAY NORMAL CURVE if you want a normal curve superimposed (see Chapter 3).

Types of Research Paper

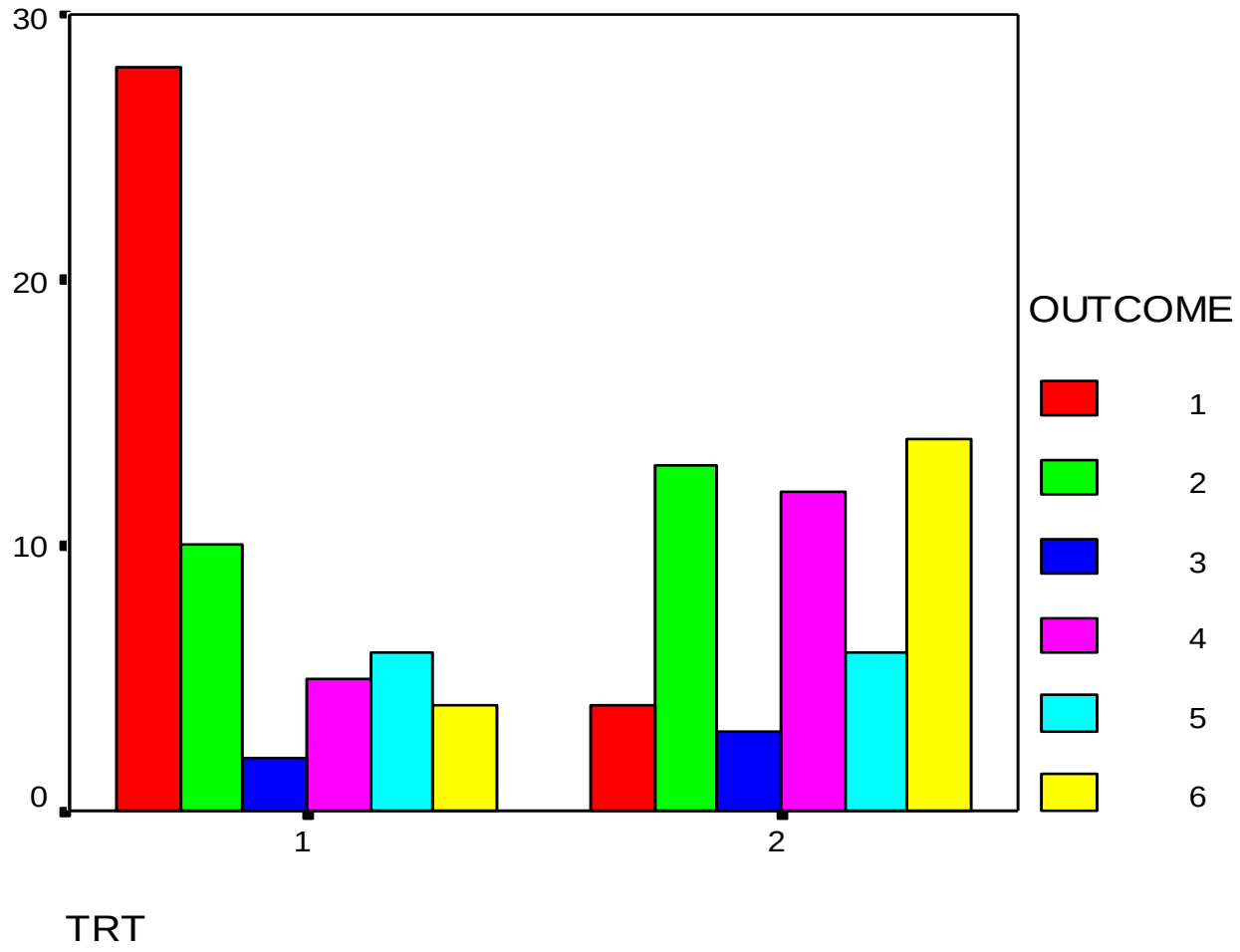
Example 1.6


30
20
10
0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
110.0
120.0
Std. Dev = 20.97
Mean = 32.1
N = 175.00
MONTHS

Side-by-Side Bar Charts

- After Importing your dataset, and providing names to variables, click on:
- GRAPHS → BAR… → Clustered (Summaries for Groups of Cases) → DEFINE
- Bars Represent N of Cases (or % of Cases)
- CATEGORY AXIS: Variable that represents groups to be compared (independent variable)
- DEFINE CLUSTERS BY: Variable that represents outcomes of interest (dependent variable)

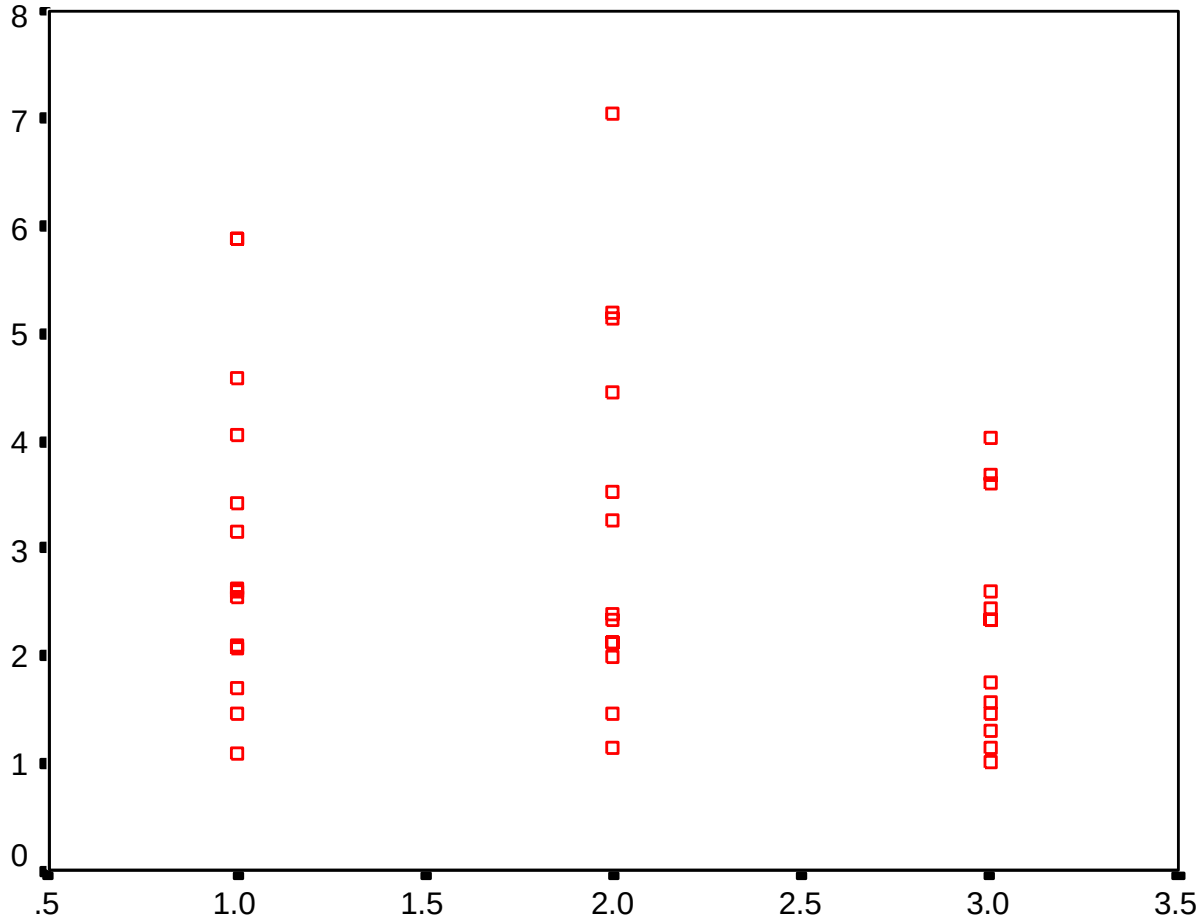
Example 1.7



Scatterplots

- After Importing your dataset, and providing names to variables, click on:
- GRAPHS → SCATTER → SIMPLE → DEFINE
- For Y-AXIS, choose the Dependent (Response) Variable
- For X-AXIS, choose the Independent (Explanatory) Variable

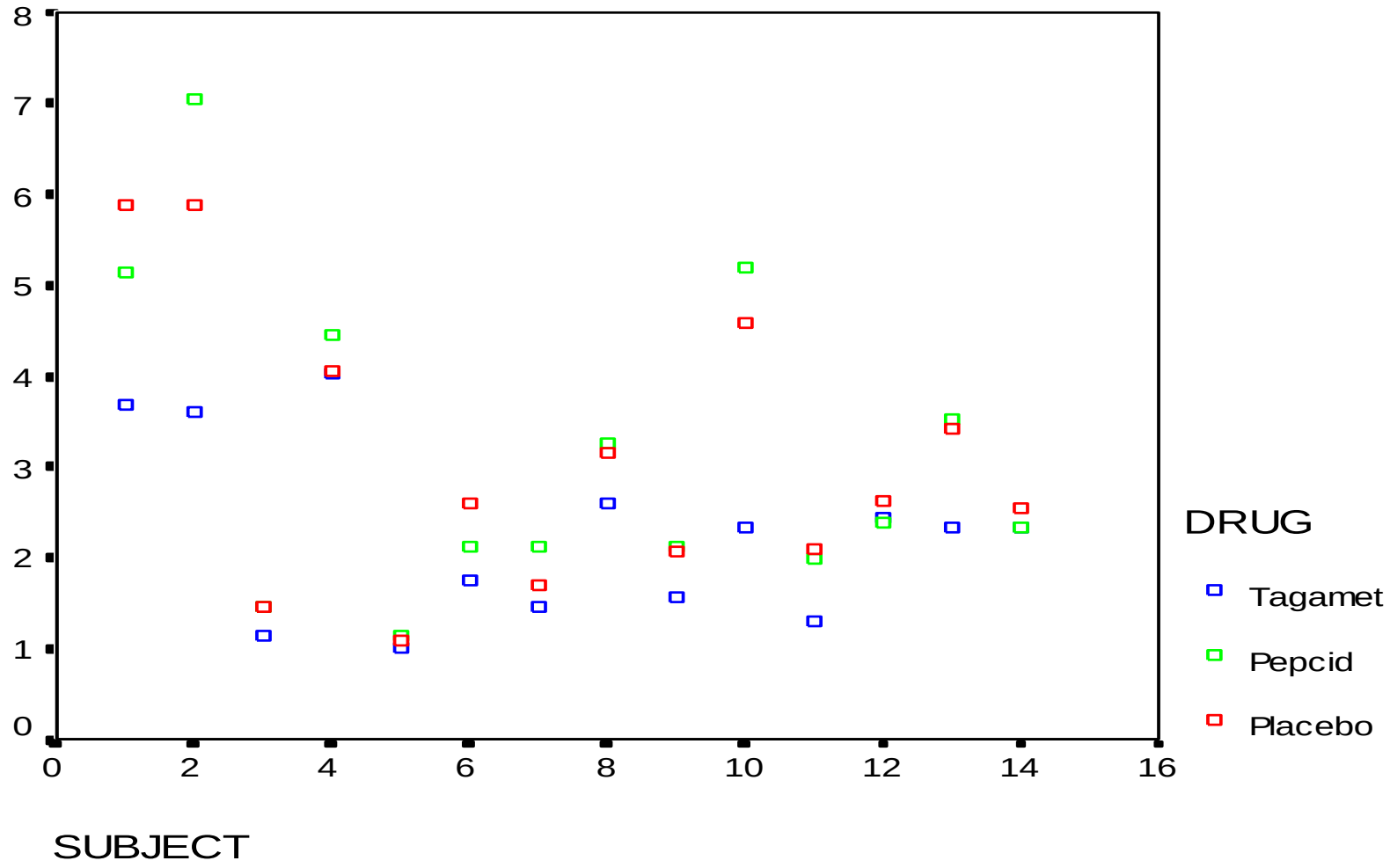
Example 1.8



Scatterplots with 2 Independent Variables

- After Importing your dataset, and providing names to variables, click on:
- GRAPHS → SCATTER → SIMPLE → DEFINE
- For Y-AXIS, choose the Dependent Variable
- For X-AXIS, choose the Independent Variable with the most levels
- For SET MARKERS BY, choose the Independent Variable with the fewest levels

Example 1.8


8
7
6
5
4
3
2
1
0
0
2
4
6
8
10
12
14
16
SUBJECT
DRUG
Tagamet
Pepcid
Placebo

Independent Sample *t*-Test

- After Importing your dataset, and providing names to variables, click on:
- ANALYZE → COMPARE MEANS → INDEPENDENT SAMPLES T-TEST
- For TEST VARIABLE, Select the dependent (response) variable(s)
- For GROUPING VARIABLE, Select the independent variable. Then define the names of the 2 levels to be compared (this can be used even when the full dataset has more than 2 levels for independent variable).

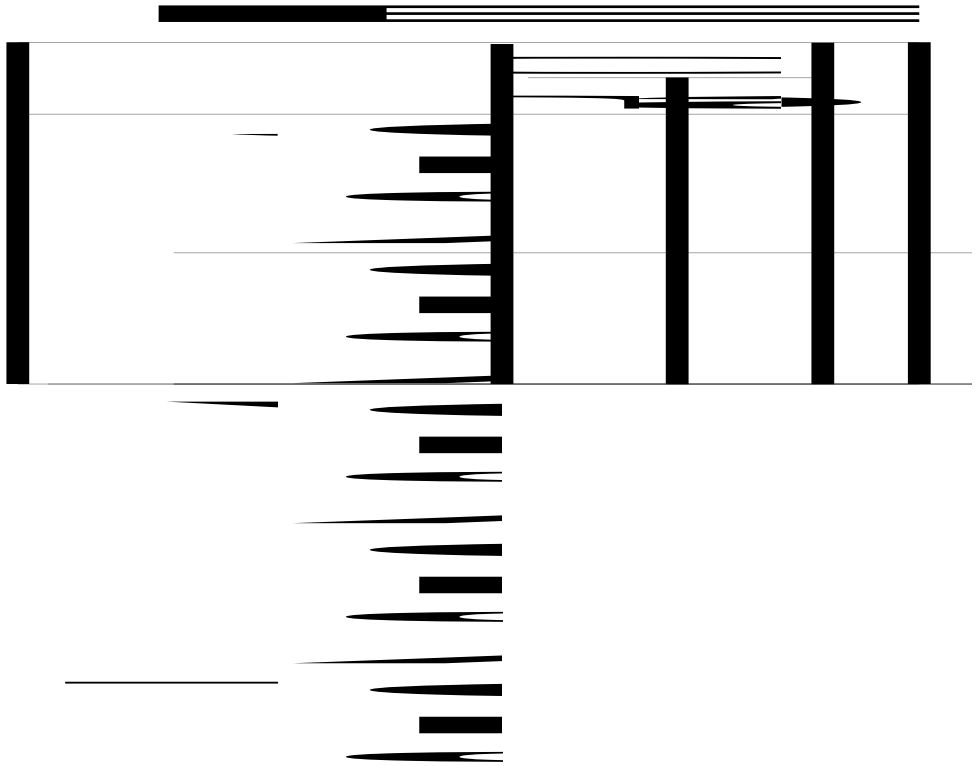
Example 3.5

Paired *t*-test

- After Importing your dataset, and providing names to variables, click on:
- ANALYZE → COMPARE MEANS → PAIRED SAMPLES T-TEST
- For PAIRED VARIABLES, Select the two dependent (response) variables (the analysis will be based on first variable minus second variable)

Chi-Square Test

- After Importing your dataset, and providing names to variables, click on:
- ANALYZE → DESCRIPTIVE STATISTICS → CROSSTABS
- For ROWS, Select the Independent Variable
- For COLUMNS, Select the Dependent Variable
- Under STATISTICS, Click on CHI-SQUARE
- Under CELLS, Click on OBSERVED, EXPECTED, ROW PERCENTAGES, and ADJUSTED STANDARDIZED RESIDUALS
- NOTE: Large ADJUSTED STANDARDIZED RESIDUALS (in absolute value) show which cells are inconsistent with null hypothesis of independence. A common rule of thumb is seeing which if any cells have values >3 in absolute value



Fisher's Exact Test

- After Importing your dataset, and providing names to variables, click on:
- ANALYZE → DESCRIPTIVE STATISTICS → CROSSTABS
- For ROWS, Select the Independent Variable
- For COLUMNS, Select the Dependent Variable
- Under STATISTICS, Click on CHI-SQUARE
- Under CELLS, Click on OBSERVED and ROW PERCENTAGES
- NOTE: You will want to code the data so that the outcome present (Success) category has the lower value (e.g. 1) and the outcome absent (Failure) category has the higher value (e.g. 2). Similar for Exposure present category (e.g. 1) and exposure absent (e.g. 2). Use Value Labels to keep output straight.

McNemar's Test

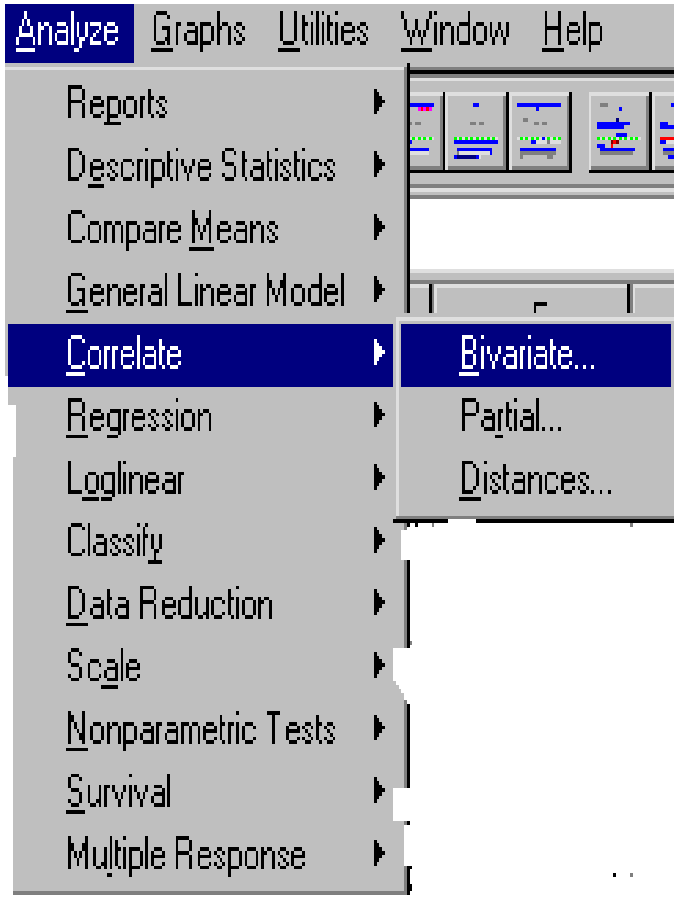
- After Importing your dataset, and providing names to variables, click on:
- ANALYZE → DESCRIPTIVE STATISTICS → CROSSTABS
- For ROWS, Select the outcome for condition/time 1
- For COLUMNS, Select the outcome for condition/time 2
- Under STATISTICS, Click on MCNEMAR
- Under CELLS, Click on OBSERVED and TOTAL PERCENTAGES
- NOTE: You will want to code the data so that the outcome present (Success) category has the lower value (e.g. 1) and the outcome absent (Failure) category has the higher value (e.g. 2). Similar for Exposure present category (e.g. 1) and exposure absent (e.g. 2). Use Value Labels to keep output straight.

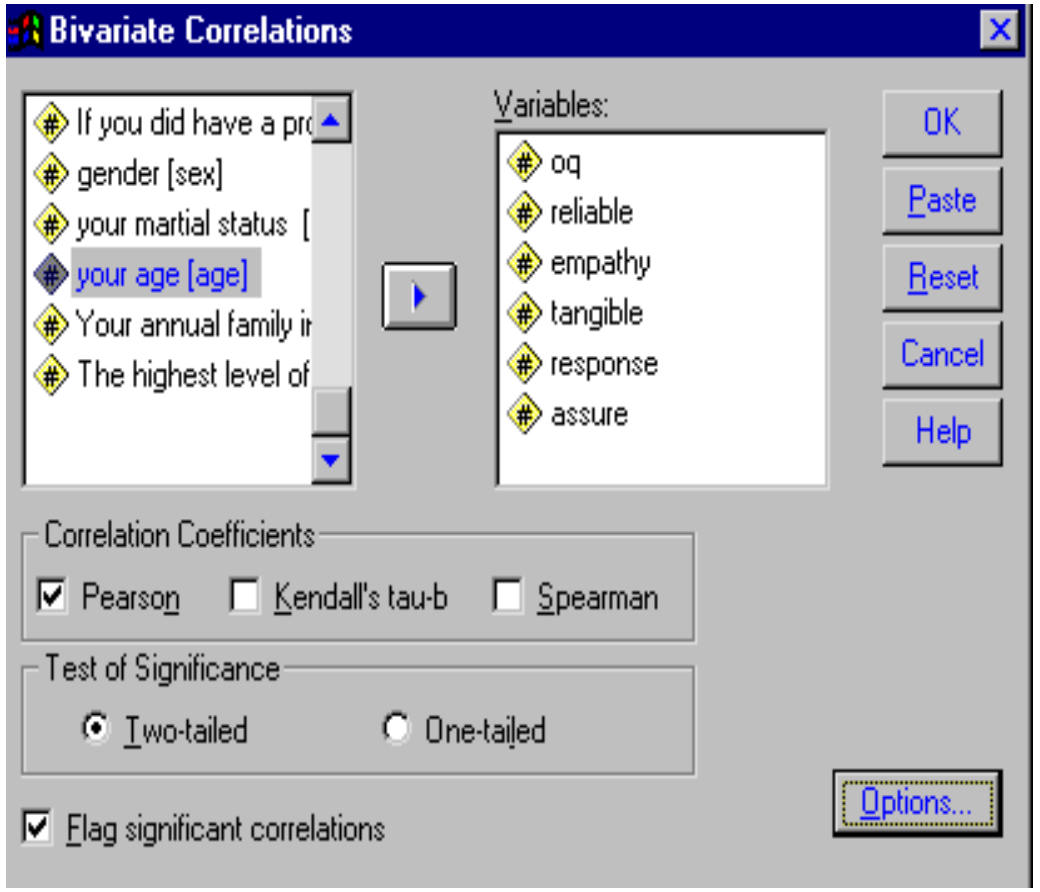
Relative Risks and Odds Ratios

- After Importing your dataset, and providing names to variables, click on:
- ANALYZE → DESCRIPTIVE STATISTICS → CROSSTABS
- For ROWS, Select the Independent Variable
- For COLUMNS, Select the Dependent Variable
- Under STATISTICS, Click on RISK
- Under CELLS, Click on OBSERVED and ROW PERCENTAGES
- NOTE: You will want to code the data so that the outcome present (Success) category has the lower value (e.g. 1) and the outcome absent (Failure) category has the higher value (e.g. 2). Similar for Exposure present category (e.g. 1) and exposure absent (e.g. 2). Use Value Labels to keep output straight.

Correlation

- After Importing your dataset, and providing names to variables, click on:
  - ANALYZE → CORRELATE→ BIVARIATE
    - Select the VARIABLES
    - Select the PEARSON CORRELATION
  - Select the Two tailed test of significance
    - Select Flag significant correlations


Analyze
Graphs
Utilities
Window
Help
Reports
Descriptive Statistics
Compare Means
General Linear Model
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Survival
Multiple Response
Bivariate...
Partial...
Distances...


Bivariate Correlations
If you did have a pr
gender [sex]
your martial status [
your age [age]
Your annual family ir
The highest level of
Variables:
oq
reliable
empathy
tangible
response
assure
OK
Paste
Reset
Cancel
Help
Correlation Coefficients
Pearson
Kendall's tau-b
Spearman
Test of Significance
Two-tailed
One-tailed
Options...
Flag significant correlations

Correlations

		OQ	RELIABLE	EMPATHY	TANGIBLE	RESPONSE	ASSURE
OQ	Pearson Correlation	1.000	.846**	.822**	.504**	.863**	.859**
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000
	N	268	251	250	217	260	254
RELIABLE	Pearson Correlation	.846**	1.000	.826**	.581**	.867**	.842**
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000
	N	251	264	256	217	249	243
EMPATHY	Pearson Correlation	.822**	.826**	1.000	.648**	.882**	.873**
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000
	N	250	256	262	216	250	243
TANGIBLE	Pearson Correlation	.504**	.581**	.648**	1.000	.607**	.567**
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000
	N	217	217	216	219	219	213
RESPONSE	Pearson Correlation	.863**	.867**	.882**	.607**	1.000	.921**
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000
	N	260	249	250	219	262	253
ASSURE	Pearson Correlation	.859**	.842**	.873**	.567**	.921**	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.
	N	254	243	243	213	253	256

Linear Regression

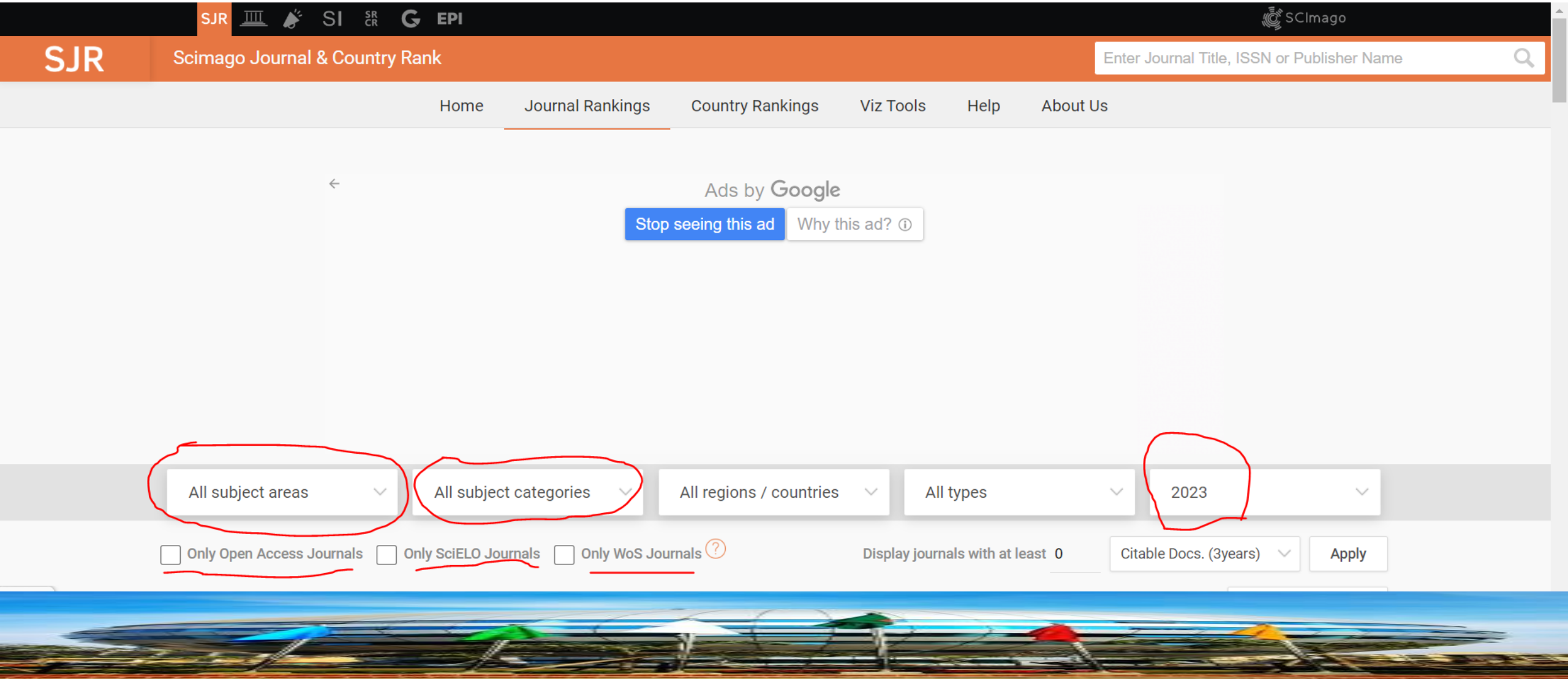
- After Importing your dataset, and providing names to variables, click on:
- ANALYZE → REGRESSION → LINEAR
- Select the DEPENDENT VARIABLE
- Select the INDEPENDENT VARAIABLE(S)
- Click on STATISTICS, then ESTIMATES, CONFIDENCE INTERVALS, MODEL FIT

Examples

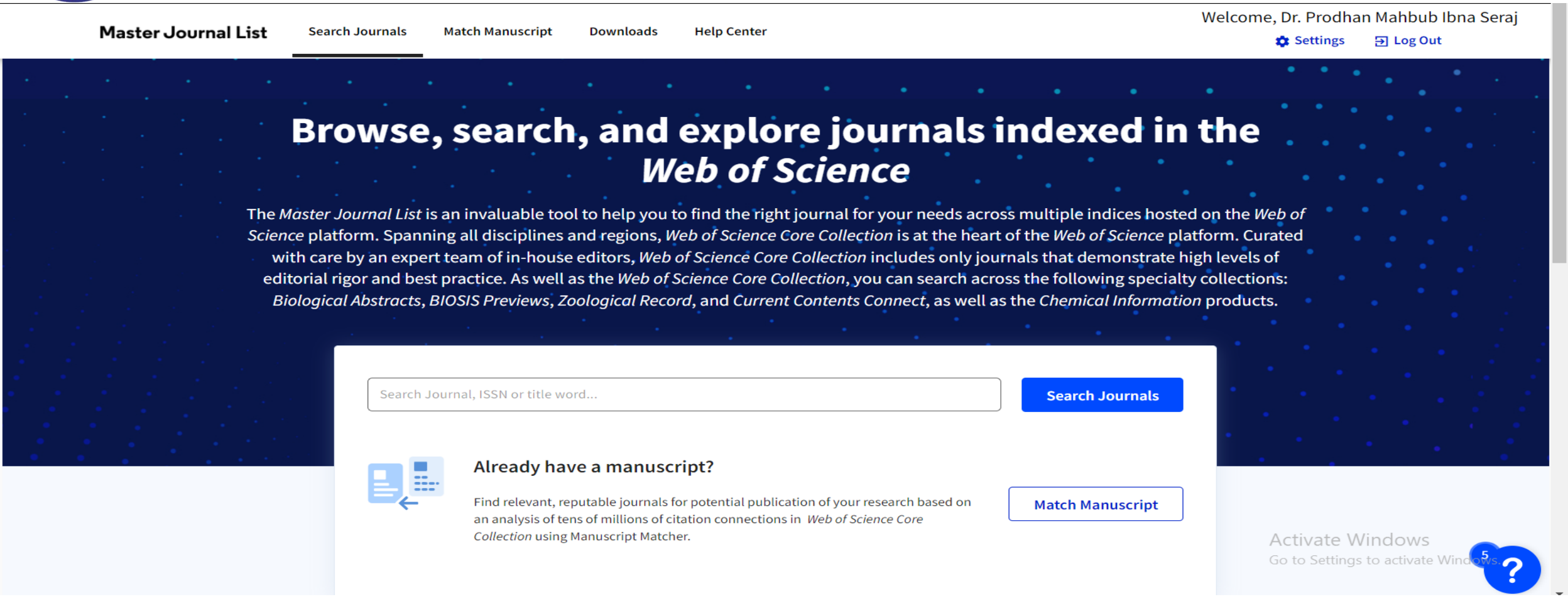
Avoid Plagiarism

- Always try to write your notes in bullet points
- Make sure that your notes are down to the bare minimum amount of the writing
- Check your paper for plagiarism and modify

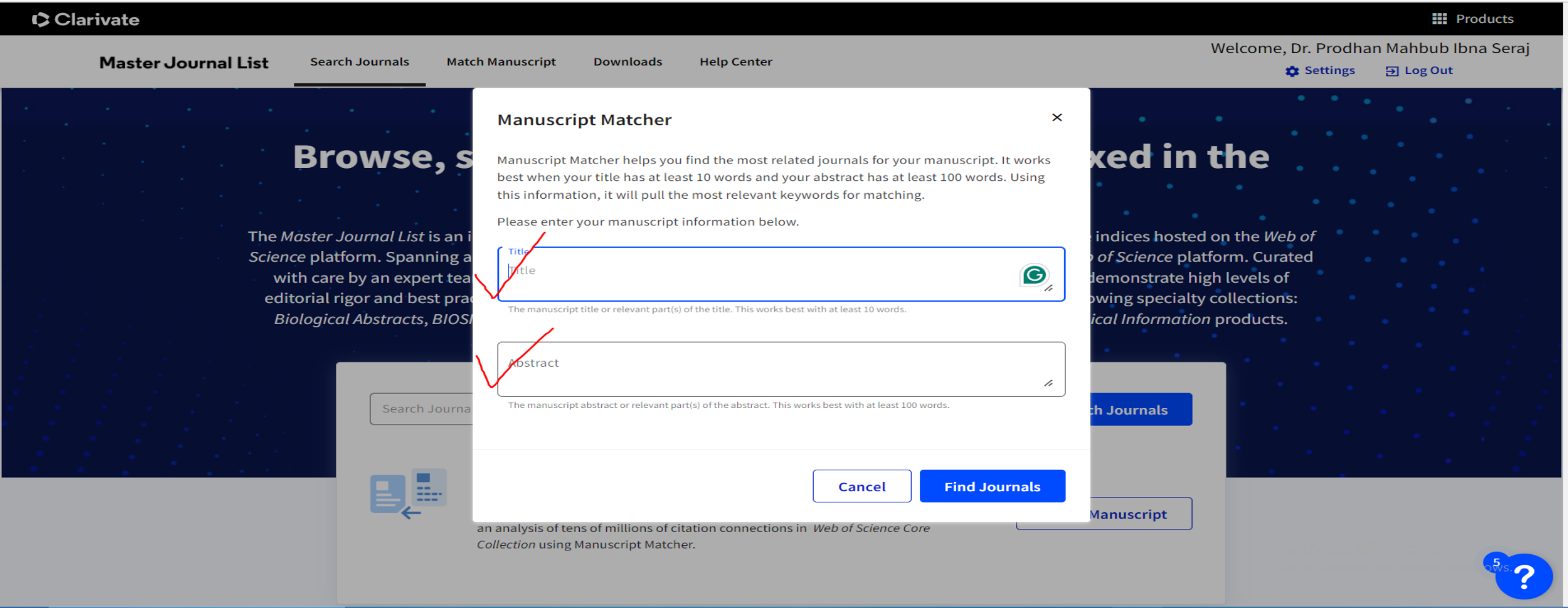
Finding Scopus Indexed Journals



Finding Scopus Indexed Journals: WOS



Finding Scopus Indexed Journals: WOS



Finding Scopus Indexed Journals: Scopus

Scopus Preview

Author Search | Sources | Create account | Sign in

Welcome to Scopus Preview

What is Scopus ↗ Blog ↗

Check access

Check if you have access through your sign in credentials or via your institution.

Check Scopus access

Check out your free author profile!

Did you know Scopus offers free profiles to all indexed authors? Review yours, claim it, and update it — all for free!

View your author profile >

Scopus content

Content coverage guide

Scopus source list

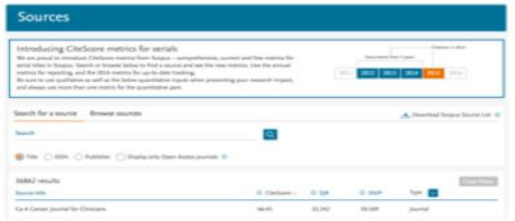
Book title list

Scopus discontinued sources list

Looking for free journal rankings and metrics?

Scopus offers free metrics to non-subscribers.

View journal rankings >



Activate Windows
Go to Settings to activate Windows.


Q&A