DegreeLive

B.Com Honours | Semester VI | Calicut University

Data Analysis Using Statistical Software

Course Code: COM6FS113 (3) • Module 2 Notes

Data Analysis Using Statistical Software
Course Code: COM6FS113 (3) • Module 2 Notes

Module II: Preparing the Data File

The credibility, precision, and reproducibility of empirical quantitative research depends fundamentally on the quality of data file preparation. In IBM SPSS Statistics, converting raw questionnaire responses, electronic forms, or financial ledgers into an analytically rigorous database requires a structured workflow encompassing metadata definition, systematic data entry, data transformation, multi-file consolidation, and comprehensive data cleaning. Errors introduced during data definition or data entry corrupt all downstream descriptive statistics, correlation matrices, and econometric models. This module examines the technical mechanics of defining variables in SPSS Variable View, entering and manipulating empirical records, importing and converting external spreadsheet datasets, merging files via cases and variables, utilizing variable sets, and executing systematic error screening and data sanitation protocols.

1. Data Architecture & Variable Definition in SPSS

In IBM SPSS Statistics, data files (.sav) do not merely hold raw numerical values; they embed an intricate metadata dictionary that defines the structural characteristics, display formatting, and analytical roles of each measurement column. All variable definition procedures occur within the **Variable View** tab of the Data Editor window.

The Eleven Pillars of SPSS Variable Architecture Variable View Protocol
Every variable created in SPSS must be parameterized across eleven standardized attribute columns: **Name**, **Type**, **Width**, **Decimals**, **Label**, **Values**, **Missing**, **Columns**, **Align**, **Measure**, and **Role**. Rigorous calibration of these eleven properties guarantees that software output tables display meaningful text labels rather than cryptic numerical codes.

1.1 Comprehensive Analysis of the Eleven Variable Properties

- **1. Name:** The unique syntax variable identifier. Must adhere strictly to SPSS naming syntax: must start with an alphabetic letter, cannot exceed 64 bytes in length, cannot contain spaces, mathematical operators (+, -, *, /), or punctuation marks, and cannot duplicate SPSS reserved syntax commands (ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH).
- **2. Type:** Specifies the data storage format. While SPSS supports several data types, four predominate in commerce research:
  - *Numeric:* The standard default format for all quantitative data, ratings, codes, and scale items. SPSS processes numeric data with maximum mathematical efficiency.
  - *String:* Text or alphanumeric characters (e.g., respondent names, open-ended qualitative comments). String variables cannot be included in mathematical or parametric statistical procedures without prior recoding into numeric codes.
  - *Date:* Standard calendar dates in diverse formats (e.g., dd-mmm-yyyy, mm/dd/yyyy). Used extensively for calculating chronological age, tenure, or holding durations via Date and Time Wizards.
  - *Custom Currency:* Numeric variables formatted with currency symbols and comma thousand separators.
- **3. Width:** Dictates the maximum number of characters or digits permitted for variable entry and storage. Default is typically 8 characters for numeric variables.

DegreeLive

- **4. Decimals:** Specifies the number of decimal places displayed on screen in Data View and output tables. For integer codes (e.g., Gender: 1, 2; Likert ratings: 1 to 5), decimals should always be set to **0** to maintain clean visual presentation. For continuous financial ratios, stock prices, and percentages, decimals are calibrated to **2** or **4**.
- **5. Label:** An extended descriptive text string (up to 256 characters) that fully identifies the variable. While the variable Name is brief (e.g., `Inv_Horizon`), the variable Label provides clear human-readable context (e.g., *"Target Investment Horizon in Years"*). SPSS automatically prints the Label on all generated frequency tables, ANOVA summaries, and charts.
- **6. Values (Value Labels):** The critical mapping mechanism that assigns meaningful verbal text descriptors to discrete numerical codes. For example, assigning `1 = Male` and `2 = Female` to a categorical variable, or `1 = Strongly Disagree` through `5 = Strongly Agree` to Likert scale items. Clicking the ellipsis (...) in the Values cell launches the Value Labels dialogue box.
- **7. Missing (User-Defined Missing Values):** Instructs SPSS how to handle non-responses, refusals, or inapplicable items. Rather than leaving blank cells (which SPSS treats as System-Missing, represented by a period '.'), researchers specify explicit discrete numerical codes:
  - `9` or `99` = Refused to Answer / Unreported
  - `8` or `98` = Not Applicable / Skipped Question
  - `-1` or `-99` = Data Corruption / Incomplete Item

  SPSS automatically filters out designated user-missing codes from statistical calculations while maintaining an exact audit count in frequency distributions.
- **8. Columns:** Controls the visual display width of the column in pixels within the Data View grid. Does not alter underlying data storage width.
- **9. Align:** Adjusts cell text alignment within Data View: Left, Right, or Centered. Numeric variables default to Right-aligned; string variables default to Left-aligned.
- **10. Measure:** Declares the Stevens' scale of measurement: **Nominal**, **Ordinal**, or **Scale** (Interval/Ratio). This declaration is vital because modern SPSS dialogue boxes and the Chart Builder use the measurement icon to automatically restrict or permit specific analytical algorithms.
- **11. Role:** Assigns the functional analytical role of the variable:
  - *Input:* Predictor or Independent Variable (IV).
  - *Target:* Outcome, Criterion, or Dependent Variable (DV).
  - *Both:* Variable serves simultaneously as input and target (e.g., mediating variables in path models).
  - *None:* Variable has no pre-assigned role; treated neutrally across all dialogue boxes.

1.2 Master Variable Architecture Reference Matrix

Variable	Type	Decimals	Variable Label	Value Labels	Missing	Measure
ID	Numeric	0	Respondent Identification Serial	None	None	Nominal
Gender	Numeric	0	Biological Gender of Investor	1 = Male, 2 = Female, 3 = Other	9	Nominal

DegreeLive

Variable	Type	Decimals	Variable Label	Value Labels	Missing	Measure
Age_Years	Numeric	0	Exact Chronological Age in Years	None	999	Scale
Education	Numeric	0	Highest Academic Level Completed	1=HSE, 2=Graduate, 3=Postgrad, 4=PhD	9	Ordinal
Monthly_Inc	Numeric	2	Monthly Net Disposable Income (INR)	None	-99	Scale
Platform	Numeric	0	Primary Trading Execution App	1=Zerodha, 2=Groww, 3=AngelOne, 4=Bank	99	Nominal
Trade_Style	Numeric	0	Primary Trading Horizon Style	1=Intraday, 2=Swing, 3=Positional, 4=F&O	9	Nominal
Risk_Pref_1	Numeric	0	"I seek high volatility for rapid gains"	1=SD, 2=D, 3=N, 4=A, 5=SA	9	Ordinal
Risk_Pref_2R	Numeric	0	"I avoid risky assets" (Reverse Coded)	1=SA(1), 2=A(2), 3=N(3), 4=D(4), 5=SD(5)	9	Ordinal
ROE_Percent	Numeric	2	Return on Equity Percentage	None	-999.00	Scale

1.3 Operational Shortcuts: Copying Variable Properties

Manually entering identical Value Labels and Missing codes across 40 or 50 Likert scale items is inefficient and error-prone. SPSS provides rapid automation shortcuts:

- **Standard Copy-Paste Method:** Complete the full definition of the first Likert variable (e.g., `Sat_1`). Click on its **Values** cell, press `Ctrl + C`, highlight the Values cells of all remaining scale variables (`Sat_2` through `Sat_50`), and press `Ctrl + V`. The complete value label dictionary is copied across all items instantaneously.
- **Define Variable Properties Utility:** Located at `Data -> Define Variable Properties`. This interactive dialogue scans existing data columns, lists all unique values encountered, and allows the analyst to enter labels, assign missing values, and change measurement levels simultaneously across dozens of variables in one window.
- **Apply Data Dictionary Wizard:** Navigating to `Data -> Copy Data Properties` launches an automated wizard that imports the complete variable definition architecture (names, labels, measurement types, missing codes) from an existing external master SPSS file (.sav) into an unformatted new dataset.

2. Data Entry & File Manipulation Operations

Once the metadata architecture is established, researchers populate the Data View grid with empirical observations. Modifying, restructuring, and derived computation represent standard operational tasks in quantitative analysis.

2.1 Structural File Modifications

- **Inserting & Deleting Cases (Rows):** To insert an omitted respondent between existing records, right-click on the row number header and select `Insert Cases`. To remove corrupted or unengaged responses, highlight the target row number and press the `Delete` key (or right-click and select `Clear`).
- **Inserting & Deleting Variables (Columns):** To introduce an omitted survey item, right-click the column header and select `Insert Variable`. A new column named `VAR00001` is created, which can be renamed in Variable View. Highlight and delete columns using `Clear`.
- **Sorting Cases:** Navigating to `Data -> Sort Cases` allows the researcher to sort observations based on one or multiple criteria. For example, sorting cases in ascending order by `Respondent_ID`, or sorting primarily by `Gender` (Ascending) and secondarily by `Annual_Income` (Descending). Sorting is an essential prerequisite before executing file merges and split-file comparisons.

2.2 Splitting Files and Selecting Cases

Splitting Datasets: Data -> Split File

Enables subgroup comparative analysis without creating separate data files:
• *Compare Groups:* Generates combined output tables placing subgroups side-by-side (e.g., comparing mean trading frequency between Male and Female investors in a single table).
• *Organize Output by Groups:* Produces separate sequential output tables for each subgroup.
• *Crucial Reset:* Remember to re-enter `Data -> Split File` and select *"Analyze all cases, do not create groups"* once subgroup analysis is finished, otherwise all subsequent procedures remain split.

Filtering Datasets: Data -> Select Cases

Isolates a specific subset of cases for targeted investigation based on logical criteria:
• *If Condition is Satisfied:* Use mathematical and logical operators (e.g., `Gender = 2 & Age_Years >= 30` to analyze only female investors aged 30 and older).
• *Filter Out Unselected Cases:* Unselected rows are temporarily crossed out in Data View (assigned a system filter variable equal to 0). Raw data is completely preserved.
• *Random Sample of Cases:* Draws an exact random percentage (e.g., 70% training set vs 30% validation set).

2.3 Weighting Cases: Data -> Weight Cases

In business research, analysts frequently receive aggregated secondary summary tables rather than individual respondent rows (e.g., a cross-tabulation showing 120 males preferred Zerodha, 80 preferred Groww, 90 females preferred Zerodha, 110 preferred Groww). Entering this in SPSS requires only 4 rows with a frequency variable `Freq_Count`, and executing `Data -> Weight Cases by Freq_Count`. SPSS mathematically treats the 4 rows as 400 discrete individual respondents in all downstream Chi-square and frequency analyses.

2.4 Transforming Variables: Recoding into Different Variables

Recoding involves transforming existing data values into new categories or numeric groupings. In empirical commerce research, researchers routinely convert continuous age into generational cohorts, or invert reverse-worded Likert statements.

Degree Live

Golden Rule of Data Transformation Methodological Safeguard
NEVER Recode into 'Same Variable': Recoding into the same variable permanently overwrites and destroys the original raw empirical data. If a coding error occurs, the original values are irretrievably lost.

ALWAYS Recode into 'Different Variables': Navigating to `Transform -> Recode into Different Variables` creates a new, separate variable column while preserving the raw source variable intact on the dataset.

```
/* SPSS Command Syntax for Recoding Continuous Age into Categorical Cohorts */
RECODE Age_Years (LOWEST THRU 25 = 1) (26 THRU 40 = 2) (41 THRU 60 = 3) (61 THRU HIGHEST = 4) (ELSE = 9)
INTO Age_Category.
VARIABLE LABELS Age_Category "Generational Investor Age Cohort".
VALUE LABELS Age_Category 1 "Gen Z (Under 25)" 2 "Millennial (26-40)" 3 "Gen X (41-60)" 4 "Senior Citizen (61+)" 9 "Missing".
EXECUTE.
```

2.5 Automatic Recode: Converting String Variables to Numeric Codes

When survey datasets contain textual responses in String format (e.g., city names: "Mumbai", "Kochi", "Bangalore", "Delhi"), parametric statistical algorithms cannot process them. Rather than manually typing numerical codes, researchers execute `Transform -> Automatic Recode`:

- SPSS automatically reads all distinct string values, sorts them alphabetically, and assigns sequential integer codes (1, 2, 3, 4).
- It creates a new numeric variable while automatically populating the **Value Labels** dictionary using the original string texts as descriptors.

2.6 Deriving Composite Scales: The Compute Variable Command

Psychometric research rarely relies on single individual survey questions. Instead, multiple Likert items measuring the same theoretical construct are aggregated into a single composite summated score using `Transform -> Compute Variable`.

- **Mean Score vs. Summated Total Score:** Researchers can compute either the arithmetic mean or the raw sum across construct items. Computing the mean preserves the original 1-to-5 Likert interpretation scale, whereas summating produces a broad continuous score.
- **Handling Missing Data during Computation:**
  - *Using Arithmetic Plus Operator (+):* If an expression is written as `COMPUTE Total_Score = Item1 + Item2 + Item3`, if a respondent skipped even one single item, SPSS assigns a System-Missing value (.) to the entire composite score.
  - *Using Statistical Functions (MEAN.n / SUM.n):* Writing `COMPUTE Mean_Score = MEAN.3(Item1, Item2, Item3, Item4)` instructs SPSS to compute the mean provided the respondent answered at least 3 out of the 4 items, mitigating unnecessary case attrition.

```
/* SPSS Command Syntax for Computing Summated Construct Score */
COMPUTE Financial_Literacy_Index = MEAN.4(FinLit_1, FinLit_2, FinLit_3, FinLit_4, FinLit_5).
VARIABLE LABELS Financial_Literacy_Index "Composite Mean Financial Literacy Index (1-5 Scale)".
VARIABLE LEVEL Financial_Literacy_Index (SCALE).
EXECUTE.
```

degreelive

3. External Data Ingestion & File Merging Operations

In contemporary business environments, primary empirical data is rarely inputted directly into SPSS keyboard-by-keyboard. Data is routinely gathered via online survey tools (Google Forms, Qualtrics, SurveyMonkey) or extracted from corporate ERP ledgers in Microsoft Excel (.xlsx) or Comma Separated Values (.csv) formats.

3.1 Importing Microsoft Excel Files into SPSS

Navigating to `File -> Open -> Data` (and setting file type filter to *Excel *.xlsx*) launches the Excel Import Wizard. Successful importation requires observing critical structural conditions:

- **First Row Variable Names:** Check the box *"Read variable names from first row of data"*. Ensure the top row of the Excel worksheet contains clean, syntax-compliant column headers without spaces or special characters.
- **Worksheet Selection:** If the Excel workbook contains multiple tabs, select the exact target sheet from the dropdown menu.
- **Range Specification:** If the Excel sheet contains decorative summary tables or notes above or beside the data, specify the exact cell coordinates (e.g., `A1:AZ500`).
- **Data Type Homogeneity:** Ensure every column in Excel contains homogeneous data types. If a numerical age column contains occasional text entries (such as "twenty-five" or "N/A"), SPSS will convert the entire column into a String variable, corrupting subsequent numerical analysis.

3.2 Merging Datasets: Adding Cases vs. Adding Variables

Large-scale research projects frequently necessitate merging disparate empirical files. SPSS provides two fundamentally distinct merging mechanisms under the `Data -> Merge Files` menu:

1. Merging Files: Add Cases (Row Concatenation)

Application: Appending additional respondents or observation periods to an existing dataset.
Scenario: Combining 250 survey responses gathered in Kerala with 250 responses gathered in Tamil Nadu.
Structural Requirement: Both datasets must possess identical variable names and measurement types. SPSS concatenates rows vertically. If variable names differ (e.g., `Income` vs `Annual_Income`), SPSS treats them as unmatched variables and creates separate columns.

2. Merging Files: Add Variables (Column Joining)

Application: Adding new survey dimensions or secondary data to the same cohort of respondents.
Scenario: Merging financial literacy survey scores with actual brokerage transaction records for the same 500 investors.
Structural Requirement: Both files must contain a common, unique identifier variable (**Key Variable**, such as `Respondent_ID` or `PAN_Hash`). **Both files must be pre-sorted in ascending order by the Key Variable** prior to executing the merge.

3.3 Managing Large Datasets: Using Variable Sets

When working with institutional corporate datasets containing hundreds of variables (e.g., annual balance sheet metrics spanning 20 years), navigating variable selection lists in dialogue boxes becomes unwieldy. SPSS provides the **Variable Sets** utility:

- **Defining Variable Sets:** Navigating to `Utilities -> Define Variable Sets` allows the researcher to group related variables into custom subsets (e.g., a set named `Liquidity_Ratios`, another named

`Demographics`, and a third named `Governance_Metrics`).

- **Activating Sets:** Navigating to `Utilities -> Use Variable Sets` allows the analyst to check only the active subsets required for current analysis. All dialogue boxes instantly collapse to show only the selected variables, dramatically accelerating workflow speed and reducing selection errors.

4. Data Screening, Quality Assurance & Data Cleaning

Data cleaning is the rigorous quality assurance process executed prior to conducting any formal hypothesis testing. Even minor data entry errors—such as accidentally typing an age of 240 instead of 24—heavily distort means, variances, and regression coefficients.

4.1 Screening Categorical Variables: Frequency Analysis

For nominal and ordinal variables (Gender, Education, Marital Status), data screening relies on checking absolute frequency distributions:

- **Procedure:** Navigate to `Analyze -> Descriptive Statistics -> Frequencies`. Move all categorical variables into the Variable box and execute.
- **Out-of-Range Error Detection:** Examine the frequency tables. If Gender is coded `1 = Male` and `2 = Female`, inspect the table for unauthorized values such as `3`, `0`, or `22` resulting from keyboard entry slip.
- **Locating and Correcting Errors:** If an invalid code '3' is detected in Gender:
  - Navigate to `Data -> Sort Cases` and sort cases by Gender to bring the offending case to the top.
  - Alternatively, use `Edit -> Find` (or press `Ctrl + F`) while highlighting the Gender column in Data View, searching for the value '3'. Cross-check the original physical paper questionnaire using the Respondent ID and correct the digital entry.

4.2 Screening Continuous Variables: Descriptive Statistics & Explore

For continuous scale variables (Age, Annual Income, Monthly Portfolio Turnover), frequency tables are inefficient because every unique number produces a separate row. Instead, researchers utilize descriptive summary statistics:

- **Procedure:** Navigate to `Analyze -> Descriptive Statistics -> Descriptives` (or the powerful `Analyze -> Descriptive Statistics -> Explore` command).
- **Minimum and Maximum Scrutiny:** In the descriptive output table, carefully inspect the **Minimum** and **Maximum** columns:
  - If `Age_Years` exhibits a Minimum of `2` or a Maximum of `450`, gross data entry errors exist.
  - If 5-point Likert items display values of `0`, `6`, or `55`, out-of-bounds errors must be flagged.
- **Mean vs. 5% Trimmed Mean:** The Explore procedure calculates both the standard Arithmetic Mean and the **5% Trimmed Mean** (which removes the top 5% and bottom 5% extreme values). If the standard mean diverges substantially from the trimmed mean, severe outliers are distorting the central tendency.

4.3 Detecting Duplicate Records

When collecting data via online electronic forms, respondents occasionally submit their survey multiple times due to slow internet connections or accidental double-clicking. SPSS provides an automated audit feature:

- **Procedure:** Navigate to `Data -> Identify Duplicate Cases`. Move identifying matching variables (e.g., Email, Phone Hash, or demographic combinations like Gender + Age + Pin Code) into the *"Define*

matching cases by" box.

- **Output Indicator:** SPSS creates a new binary system variable named `PrimaryLast` or `DuplicateIndicator` (where `0 = Duplicate Case` and `1 = Primary Unique Case`). Researchers can instantly filter or delete all '0' entries using `Data -> Select Cases`.

4.4 Systematic Missing Data Protocols

4.5 Univariate Outlier Detection via Standardized Z-Scores

An outlier is an observation that deviates markedly from the overall distribution of data. In continuous financial and behavioral measures, undetected outliers distort the sample mean and inflate standard deviations, creating false rejections or acceptances of hypotheses:

- **Standardized Z-Score Calculation:** Navigating to `Analyze -> Descriptive Statistics -> Descriptives` and checking the option *"Save standardized values as variables"* instructs SPSS to calculate and append a new Z-score column (e.g., `ZAge`, `ZIncome`) to the active dataset.
- **Z-Score Evaluation Benchmarks:**
  - For moderate sample sizes (N < 100): Z-scores exceeding ± 2.58 (p < 0.01) indicate potential univariate outliers.
  - For large sample sizes (N > 100): In accordance with Tabachnick & Fidell (2019), cases exhibiting **Z-scores exceeding ± 3.29 (p < 0.001)** are flagged as severe univariate outliers.
- **Treatment Protocols:** Analysts must decide whether to (1) verify the original source for typographical entry error, (2) delete the outlier case if it represents an unrepresentative population element, or (3) apply *Winsorization* (re-coding the extreme value to the next highest non-outlying score at the 99th percentile).

4.6 Multivariate Outlier Detection: Mahalanobis Distance

A case may not be an outlier on any single isolated variable, but may represent a severe multidimensional outlier when combinations of variables are examined concurrently (e.g., a respondent reporting an Age of 19 years combined with an Annual Income of ₹50 Lakhs and 15 years of trading experience). Multivariate outliers distort multivariate analyses (Regression, MANOVA, Factor Analysis):

- **Procedure via Linear Regression:** Navigate to `Analyze -> Regression -> Linear`. Specify an arbitrary case identifier as the Dependent variable, enter all continuous survey variables as Independents, click `Save`, and check **Mahalanobis**.
- **Interpretation:** SPSS computes a new variable named `MAH_1` representing the Mahalanobis Distance squared (D-squared) of each case from the multivariate centroid.
- **Chi-Square Significance Criterion:** Using a Chi-Square distribution lookup table at alpha = 0.001 with degrees of freedom equal to the number of predictor variables: any case where the computed Mahalanobis distance exceeds the critical Chi-Square threshold is flagged as a significant multivariate outlier and evaluated for exclusion.

4.7 Missing Data Mechanism Classification: Little's MCAR Test

Understanding the underlying statistical mechanism driving missing data determines which treatment protocol is methodologically valid:

- **Missing Completely at Random (MCAR):** Missingness is totally unrelated to any observed or unobserved variables (e.g., a questionnaire page was accidentally dropped by a scanner). The remaining complete cases represent an unbiased random subsample.

degreelive

- **Missing at Random (MAR):** Missingness is related to an observed variable in the dataset, but not to the value of the missing item itself (e.g., older investors are less likely to report online trading frequency, but conditional on age, missingness is random).
- **Missing Not at Random (MNAR):** Missingness depends directly on the unobserved value itself (e.g., ultra-high-income individuals refusing to disclose income due to tax avoidance sensitivity). Creates severe systematic non-response bias.
- **Little's MCAR Test:** Available under `Analyze -> Missing Value Analysis`. Evaluates the null hypothesis that data is Missing Completely at Random. If the test statistic yields **$p > 0.05$**, the null hypothesis is retained, confirming the data is MCAR and validating listwise deletion or simple imputation. If **$p < 0.05$**, data is not MCAR, and advanced techniques such as Multiple Imputation (MI) or Full Information Maximum Likelihood (FIML) must be deployed.

Summary Synthesis

Missing Handling Method	Operational Mechanism	Methodological Implications
Listwise Deletion (Complete Case Analysis)	Excludes a respondent entirely from the analysis if they have a missing value on *any single variable* included in the model.	Guarantees that all statistical parameters are estimated on the exact same cohort, but causes massive statistical power loss if missingness is widespread across many items.
Pairwise Deletion	Excludes a respondent only from specific bivariate correlations where one of the two variables is missing.	Preserves total sample size across correlation matrices, but produces mathematically inconsistent sample sizes across different cells of the correlation table.
Mean Substitution / Imputation	Replaces missing data points with the overall mean of the variable across valid cases (`Transform -> Replace Missing Values`).	

Operational Phase	Key SPSS Commands & Tools	Quality Assurance Objective
Variable Definition	Variable View (Name, Type, Label, Values, Missing, Measure).	Embeds a comprehensive metadata dictionary, ensuring self-documenting analytical outputs.
Data Transformation	Recode into Different Variables, Compute Variable (MEAN.n).	Generates composite construct indices while safeguarding raw empirical data from destruction.
Data Ingestion	Excel Import Wizard, Delimited Text Ingestion.	Bridges spreadsheet platforms and statistical databases with zero structural truncation.
File Consolidation	Merge Files: Add Cases (Rows) vs Add Variables (Columns via Key).	Integrates multi-source primary and secondary datasets into a unified analytical matrix.

Operational Phase	Key SPSS Commands & Tools	Quality Assurance Objective
Quality Screening	Frequencies (Out-of-range codes), Descriptives (Min/Max checks), Identify Duplicates.	Purges data entry errors, eliminating bias prior to executing inferential hypothesis tests.

DegreeLive

Acing your exams is just a click away!
Visit www.degreelive.in to download the next module for free.

DegreeLive