DegreeLive

B.Com Honours | Semester VI | Calicut University

# Data Analysis Using Statistical Software

Course Code: COM6FS113 (3) • Module 4 Notes

Data Analysis Using Statistical Software
Course Code: COM6FS113 (3) • Module 4 Notes

# Module IV: Overview of Different Statistical Techniques

Once raw empirical data has been systematically prepared, screened for distributional normality, diagnosed for outliers, and validated for internal measurement reliability, the econometrician or business analyst proceeds to substantive inferential hypothesis testing. Inferential statistical techniques enable researchers to draw conclusions regarding population parameters based on observed sample statistics. In modern commerce, finance, and management investigations, analytical techniques fall into three broad scientific categories: (1) **Techniques of Association and Prediction** (Pearson, Spearman, and Partial Correlation; Multiple Linear Regression); (2) **Techniques of Data Reduction and Dimension Exploration** (Exploratory Factor Analysis and Principal Component Analysis); and (3) **Techniques of Group Mean Comparison** (Independent Samples T-Test, Paired Samples T-Test, One-Way ANOVA, Two-Way ANOVA, MANOVA, and ANCOVA). This module provides an exhaustive, textbook-depth exposition of the theoretical assumptions, SPSS execution workflows, output interpretation protocols, and academic reporting conventions governing these statistical methodologies.

## 1. Correlation Analysis: Bivariate and Partial Association

Correlation analysis evaluates the direction, strength, and statistical significance of the association between two or more metric variables without implying a cause-and-effect relationship.

Taxonomy of Correlation Techniques Analytical Selection
**1. Pearson Product-Moment Correlation (r):** Parametric technique measuring the linear relationship between two continuous variables (Interval/Ratio scale). Assumes bivariate normality, linearity, and homoscedasticity.

**2. Spearman Rank-Order Correlation (rho) / Kendall's Tau-b:** Non-parametric technique measuring monotonic association between two ordinal variables or continuous variables violating parametric normality assumptions.

**3. Partial Correlation (r_xy.z):** Isolates and quantifies the net direct correlation between two continuous variables while statistically controlling for (partialling out) the confounding influence of one or more third variables.

### 1.1 Pearson Product-Moment Correlation

The Pearson correlation coefficient (denoted as *r*) ranges strictly between -1.00 and +1.00:

- **Positive Correlation (0.00 < r ≤ 1.00):** Both variables move in the same direction; as one variable increases, the other systematically increases (e.g., Advertising Expenditure and Sales Revenue).
- **Negative / Inverse Correlation (-1.00 ≤ r < 0.00):** Variables move in opposite directions; as one variable increases, the other systematically decreases (e.g., Interest Rates and Bond Prices).
- **Zero Correlation (r = 0.00):** Indicates the complete absence of any linear association between the two variables.

**Pearson Correlation Coefficient Formula:**
r = Covariance(X, Y) / [Standard Deviation(X) * Standard Deviation(Y)]

**Standard Empirical Strength Benchmarks (Cohen, 1988):**
• Small / Weak Correlation: |r| = 0.10 to 0.29
• Medium / Moderate Correlation: |r| = 0.30 to 0.49
• Large / Strong Correlation: |r| = 0.50 to 1.00
**Coefficient of Determination (r-squared):** Represents the proportion of shared variance between the two variables (e.g., r = 0.60 implies r-squared = 0.36, meaning 36% of the variance is shared).

### 1.2 Spearman Rank-Order Correlation

When data is measured on an ordinal scale (e.g., customer satisfaction ranks, credit risk ratings) or when continuous data severely violates normality, researchers deploy **Spearman's Rho**. Spearman converts raw scores into ranks from 1 to N and calculates the Pearson correlation on those ranks. It evaluates whether the relationship is *monotonic* (whether variables move together in direction) without requiring linear constancy.

### 1.3 Partial Correlation: Controlling for Confounding Variables

In business research, a significant bivariate correlation between two variables (X and Y) may be spurious, driven entirely by a common third variable (Z). For example, a high positive correlation between Financial Knowledge (X) and Investment Wealth (Y) might simply be driven by Age (Z), as older individuals naturally accumulate both experience and capital.

Partial correlation isolates the unique association between X and Y by mathematically holding Z constant across all participants. Executed in SPSS via `Analyze -> Correlate -> Partial`, entering X and Y in the *Variables* box and Z in the *Controlling for* box.

```
/* SPSS Syntax: Bivariate & Partial Correlation */
CORRELATIONS
/VARIABLES=Financial_Literacy Investment_Volume Age
/PRINT=TWOTAIL NOSIG.

PARTIAL CORR
/VARIABLES=Financial_Literacy Investment_Volume BY Age
/SIGNIFICANCE=TWOTAIL.
```

## 2. Multiple Linear Regression Analysis

Multiple linear regression is an advanced parametric modeling technique that predicts the value of a continuous dependent (criterion) variable (Y) based on the linear combination of two or more continuous or dummy-coded independent (predictor) variables (X1, X2, ..., Xk).

**Population Multiple Regression Model:**
Y = Beta_0 + Beta_1 * X1 + Beta_2 * X2 + ... + Beta_k * Xk + Error

**Sample Estimated Regression Equation:**
Y_hat = b0 + b1 * X1 + b2 * X2 + ... + bk * Xk

**Where:**
• b0 = Intercept (the expected value of Y when all X predictors equal zero).
• b1, b2, ..., bk = Partial unstandardized regression slopes (the expected change in Y for a one-unit change in X_k, holding all other predictors constant).
• Beta_k = Standardized coefficients (measuring the relative predictive impact of each predictor in standard deviation units).

DegreeLive

## 2.1 Regression Model Building Strategies

SPSS supports multiple distinct entry methods for incorporating predictor variables into the regression equation:

- **Standard / Simultaneous Regression (Enter Method):** All independent variables are entered into the equation simultaneously. Each predictor is evaluated in terms of what it adds to the prediction of the dependent variable beyond all other predictors. This is the gold standard when theory does not specify a causal order.
- **Hierarchical / Sequential Blockwise Regression:** The researcher specifies the exact order in which variables enter the model based on prior theoretical logic or temporal sequence. For example, demographic control variables (Age, Gender, Education) are entered in Block 1, followed by substantive psychological variables (Risk Aversion, Financial Knowledge) in Block 2. Researchers evaluate the **R-Square Change (Delta R^2)** and **F-Change significance** to determine whether Block 2 accounts for statistically significant incremental variance beyond the baseline control variables.
- **Statistical / Stepwise Regression (Forward, Backward, Stepwise):** Variable entry and removal decisions are made purely mathematically based on algorithmic statistical criteria (probability of F-to-enter ≤ 0.05, probability of F-to-remove ≥ 0.10). While popular in exploratory predictive data mining, methodologists caution against stepwise methods in confirmatory theory-testing due to sample-specific capitalization on chance.

## 2.2 Dummy Variable Coding for Categorical Predictors

Multiple regression mathematically requires all independent variables to be continuous (interval or ratio). However, business investigations routinely explore categorical predictors such as Gender, Geographic Region, Industry Sector, or Credit Rating. To integrate categorical predictors into regression, researchers convert them into **Dummy Variables** (dichotomous variables coded 0 or 1):

- **The (k - 1) Dummy Rule:** For any categorical variable containing *k* discrete categories, exactly **k - 1 dummy variables** must be constructed and included in the regression model. Including all *k* dummy variables alongside the intercept triggers the *"Dummy Variable Trap"* (perfect multicollinearity, causing mathematical singularity).
- **Reference / Baseline Category:** The single omitted category serves as the benchmark reference group against which all dummy regression coefficients are interpreted.
- **Interpretation of Dummy Slope (B):** The unstandardized coefficient *B* for a dummy variable represents the estimated mean difference on the dependent variable between that specific dummy group and the omitted reference group, holding all other predictors constant.

```
/* SPSS Syntax: Constructing Dummy Variables for a 3-Category Region Variable */
/* Reference Category = North (Region = 1) */
COMPUTE Dummy_South = (Region = 2).
COMPUTE Dummy_East = (Region = 3).
EXECUTE.

/* Running Regression with Dummy Predictors */
REGRESSION
/DEPENDENT Sales_Volume
/METHOD=ENTER Advertising_Exp Dummy_South Dummy_East.
```

## 2.3 SPSS Execution Protocol

- **Step 1:** Navigate to `Analyze -> Regression -> Linear`.

- **Step 2:** Assign the continuous outcome variable into the **Dependent** box.
- **Step 3:** Assign the predictor variables into the **Independent(s)** box. Choose the Method: *Enter* (simultaneous standard regression) or *Stepwise / Hierarchical*.
- **Step 4 (Statistics):** Click `Statistics`. Select *Estimates*, *Confidence intervals*, *Model fit*, *R squared change*, *Descriptives*, **Collinearity diagnostics**, and **Durbin-Watson**.
- **Step 5 (Plots):** Click `Plots`. Assign `*ZRESID` to the Y-axis and `*ZPRED` to the X-axis (for testing homoscedasticity). Check **Histogram** and **Normal probability plot** (for residual normality).

## 2.4 Deciphering the Core Regression Output Tables

SPSS produces four central tables that must be interpreted sequentially:

- **1. Model Summary Table:**
  - *R (Multiple Correlation Coefficient):* The correlation between the observed Y and the values predicted by the model.
  - *R Square (Coefficient of Determination):* The proportion of total variance in the dependent variable explained by the combined set of predictors (e.g., R-Square = 0.542 indicates 54.2% of the variance in Investment Volume is explained by the model).
  - *Adjusted R Square:* Corrects R Square for sample size and the number of predictors. Prevents artificial inflation from adding irrelevant variables. In publishable research, Adjusted R Square is the standard reported goodness-of-fit metric.
  - *Durbin-Watson Statistic:* Tests the assumption of independent errors (absence of autocorrelation in residuals). Values range from 0 to 4. A value close to **2.00** (acceptable range: 1.50 to 2.50) confirms independence of residuals. Values < 1.50 indicate positive autocorrelation.
- **2. ANOVA Table (F-Test of Overall Model Significance):**
  - Evaluates the null hypothesis that all population slope coefficients simultaneously equal zero (H0: Beta_1 = Beta_2 = ... = Beta_k = 0).
  - Reports Regression Sum of Squares, Residual Sum of Squares, Degrees of Freedom, Mean Squares, the **F-statistic**, and the p-value (**Sig.**). If **p < 0.05**, the overall regression model is statistically significant, proving that the predictors collectively explain a significant proportion of variance.
- **3. Coefficients Table (Individual Predictor Parameter Estimates):**
  - *Unstandardized Coefficients (B and Std. Error):* Used to construct the raw predictive regression equation.
  - *Standardized Coefficients (Beta):* Converted to standard deviation metrics, allowing direct comparison of predictor importance. A predictor with Beta = 0.45 exerts three times greater relative impact on the dependent variable than a predictor with Beta = 0.15.
  - *t-statistic and Sig. (p-value):* Tests whether each individual predictor makes a unique, statistically significant contribution to the model (p < 0.05 rejects H0: b_k = 0).
- **4. Collinearity Diagnostics Table:**
  - *Tolerance:* Calculated as (1 - R_i^2), where R_i^2 is the variance of predictor X_i explained by all other predictors. **Tolerance < 0.10 indicates severe multicollinearity**.
  - *Variance Inflation Factor (VIF):* The inverse of Tolerance (VIF = 1 / Tolerance). **VIF > 10** indicates severe multicollinearity that artificially inflates standard errors, making regression coefficients unstable. Values of VIF < 3.0 or 5.0 are considered safe.

DegreeLive

### 2.5 Detailed Empirical Regression Case Illustration

Consider an empirical study examining corporate financial performance. A financial econometrician models Return on Equity (ROE) using three predictors: Financial Leverage (Debt-to-Equity Ratio), Firm Size (Natural Log of Total Assets), and Liquidity (Current Ratio) across N = 200 publicly listed commercial firms:

| Model Component | B (Unstandardized) | Std. Error | Beta (Std.) | t-statistic | p-value (Sig.) |
|---|---|---|---|---|---|
| **(Constant / Intercept)** | 4.120 | 1.085 | -- | 3.797 | 0.000 |
| **Financial Leverage** | -1.840 | 0.420 | -0.312 | -4.381 | 0.000 |
| **Firm Size (Log Assets)** | 2.150 | 0.310 | 0.485 | 6.935 | 0.000 |
| **Current Ratio** | 0.650 | 0.280 | 0.148 | 2.321 | 0.021 |

**Model Diagnostic Summary:** R = 0.642, R-Square = 0.412, Adjusted R-Square = 0.403, $F(3, 196) = 45.73$, $p < 0.001$. Durbin-Watson = 1.94 (confirming independence of residuals). All VIF values fall between 1.21 and 1.48 (confirming complete absence of multicollinearity).

**Academic Interpretation:** The overall regression model accounts for 40.3% of the variance in corporate Return on Equity. Firm Size is the most influential positive driver of profitability (Beta = 0.485, $p < 0.001$), indicating substantial economies of scale. Financial Leverage exerts a significant negative influence (Beta = -0.312, $p < 0.001$), reflecting the burden of borrowing costs on equity returns. Current Ratio exerts a modest but statistically significant positive effect (Beta = 0.148, $p = 0.021$).

## 3. Exploratory Factor Analysis (EFA) & Principal Component Analysis (PCA)

Exploratory Factor Analysis (EFA) is a multivariate data-reduction and dimension-exploration technique. In commerce, marketing, and organizational research, researchers often administer survey instruments containing dozens of individual items. EFA analyzes the inter-item correlation matrix to identify a smaller set of underlying, unobserved latent dimensions (factors) that account for the shared variance among the items.

EFA Procedural Flowchart Methodological Steps

**Step 1: Suitability Assessment & Factorability:** Compute KMO (> 0.60) and Bartlett's Test of Sphericity ($p < 0.001$).

**Step 2: Factor Extraction:** Select Principal Component Analysis (PCA) or Principal Axis Factoring (PAF).

**Step 3: Determining Number of Factors:** Kaiser Rule (Eigenvalue > 1.0) and Scree Plot visual inspection.

**Step 4: Factor Rotation:** Orthogonal (Varimax) for independent factors or Oblique (Direct Oblimin) for correlated factors.

**Step 5: Interpretation & Naming:** Examine Rotated Component Matrix, assess factor loadings (> 0.40 / 0.50), evaluate communalities, and assign theoretical construct names.

**Step 6: Saving Factor Scores:** Generate standardized composite factor scores for subsequent regression or ANOVA modeling.

### 3.1 Evaluating Data Suitability: KMO & Bartlett's Test

Prior to extracting factors, the correlation matrix must be tested to ensure sufficient correlation structure exists:

- **Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy:** Quantifies the proportion of variance among variables that might be common variance. Ranges from 0.00 to 1.00:
  - *KMO ≥ 0.90:* Marvelous / Superb.
  - *0.80 ≤ KMO < 0.90:* Meritorious / Great.
  - *0.70 ≤ KMO < 0.80:* Middling / Good.
  - *0.60 ≤ KMO < 0.70:* Mediocre (bare minimum acceptable threshold).
  - *KMO < 0.60:* Unacceptable; factor analysis is inappropriate.
- **Bartlett's Test of Sphericity:** Tests the null hypothesis that the correlation matrix is an **identity matrix** (a matrix where all diagonal elements equal 1.0 and all off-diagonal correlations equal 0.0). Researchers require a **statistically significant result (p < 0.001)** to reject the null hypothesis, confirming that meaningful inter-item correlations exist to warrant factor extraction.

### 3.2 Extraction Methods: PCA vs Common Factor Analysis

Researchers must select the appropriate mathematical extraction method in the *Extraction* sub-dialog:

- **Principal Component Analysis (PCA):** Analyzes the *total variance* (common variance + specific variance + error variance) present in the variables. It mathematically transforms correlated items into orthogonal linear composites. PCA is the preferred method when the primary research objective is purely empirical data reduction (reducing 25 questionnaire items into 4 manageable summary components).
- **Principal Axis Factoring (PAF / Common Factor Analysis):** Analyzes strictly the *shared / common variance* among items, replacing the 1.0s on the diagonal of the correlation matrix with communality estimates (squared multiple correlations). PAF is the method of choice when the primary objective is latent construct modeling and uncovering underlying psychological traits.

### 3.3 Factor Retention Decision Rules

- **1. Kaiser's Criterion (Eigenvalues-Greater-Than-One Rule):** An eigenvalue represents the total amount of variance accounted for by a given factor. By default, SPSS retains all factors exhibiting an **Eigenvalue ≥ 1.00**. In PCA, since each standardized variable contributes a variance of 1.00, any factor with an eigenvalue below 1.0 explains less variance than a single original variable and is discarded.
- **2. Cattell's Scree Plot:** A two-dimensional graph plotting factor eigenvalues against factor numbers. The curve slopes steeply downward and then levels off into a horizontal plateau ("scree"). The researcher identifies the **"elbow" or point of inflexion** and retains only the factors located above the elbow.
- **3. Cumulative Percentage of Variance Explained:** In humanities and social sciences, retained factors should account for at least **50% to 60%** of the total cumulative variance across all items.

### 3.4 Factor Rotation & Matrix Interpretation

Unrotated factor matrices are mathematically optimal but conceptually difficult to interpret because the first factor typically loads heavily on nearly all items. Factor rotation repositions the axes in multidimensional space to achieve **Simple Structure (Thurstone, 1947)**, where each variable loads strongly on only one factor and near zero on all others.

degreelive

| Rotation Strategy | Underlying Theoretical Assumption | Primary Algorithms & Output |
|---|---|---|
| **Orthogonal Rotation** | Assumes that the underlying latent factors are completely uncorrelated and orthogonal (90-degree angle) to one another. | **Varimax:** Maximizes variance of squared loadings, sharpening factor distinctions. Produces the *Rotated Component Matrix*. Most widely used in business research. |
| **Oblique Rotation** | Assumes that the underlying latent dimensions are realistically correlated in natural human behavior and business settings. | **Direct Oblimin & Promax:** Allows factor axes to rotate freely at angles other than 90 degrees. Produces the *Pattern Matrix*, *Structure Matrix*, and *Factor Correlation Matrix*. |

In the resulting **Rotated Component Matrix**, researchers examine factor loadings:
• *Factor Loading:* The correlation between an item and a factor. Loadings ≥ 0.40 are considered meaningful, and loadings ≥ 0.50 are considered practically significant (Hair et al., 2019).
• *Cross-Loading:* An item that loads ≥ 0.40 on two or more distinct factors is a cross-loading item and is typically deleted to preserve clean factor structures.
• *Communalities (h^2):* The proportion of an item's variance explained by the extracted factors. Items with communalities < 0.40 or 0.50 are weak and should be considered for removal.

### 3.5 Saving and Utilizing Factor Scores

In SPSS, clicking the `Scores` button and checking **"Save as variables"** allows researchers to compute individual standardized factor scores for every respondent in the dataset. SPSS offers three computation methods:

- **Regression Method:** The default procedure; calculates factor score coefficients that maximize correlation between scores and factors, though scores can be slightly correlated even under orthogonal rotation.
- **Bartlett Method:** Produces unbiased factor scores that correlate only with their corresponding factor.
- **Anderson-Rubin Method:** An adjustment to the Bartlett method that guarantees that saved factor scores are completely uncorrelated with one another (mean of 0, standard deviation of 1). Highly recommended when saved factor scores will serve as independent predictor variables in a subsequent multiple regression model, completely eliminating multicollinearity.

## 4. Comparison of Means: T-Tests and Analysis of Variance

When business researchers wish to investigate whether mean differences across experimental conditions or demographic cohorts are statistically significant, they employ mean comparison tests.

### 4.1 One-Sample T-Test

Compares the observed sample mean of a single continuous variable against a known or hypothetical population benchmark (Test Value). For example, testing whether the average audit completion time across commercial banks differs significantly from a regulatory standard benchmark of 14.0 days.

### 4.2 Independent Samples T-Test

Used to compare the means of a continuous dependent variable between **two independent, unrelated groups** (e.g., Comparing mean trading returns between Male and Female investors, or comparing job satisfaction between Remote and In-Office employees).

DegreeLive

- **Critical Parametric Assumption:** Homogeneity of Variances (the assumption that the variance of the dependent variable is equal across both groups).
- **Levene's Test for Equality of Variances:** SPSS outputs Levene's F-test alongside the t-test:
  - *p > 0.05:* Fail to reject Ho → Variances are equal → Read the top row: **"Equal variances assumed"**.
  - *p ≤ 0.05:* Reject Ho → Variances are unequal → Read the bottom row: **"Equal variances not assumed" (Welch's t-test)**, which adjusts the degrees of freedom downward to protect against Type I error.
- **Interpreting Significance:** If the 2-tailed p-value (Sig.) is < 0.05, the difference between the two group means is statistically significant.
- **Effect Size (Cohen's d):** Calculated as (Mean_1 - Mean_2) / Pooled SD. Benchmarks: d = 0.20 (small), d = 0.50 (medium), d = 0.80 (large effect).

### 4.3 Paired Samples T-Test (Dependent / Repeated Measures)

Used to compare means from the **same group of respondents measured on two different occasions or under two different conditions** (e.g., Evaluating Financial Knowledge scores *Before Training* vs *After Training*). Calculates the difference score for each participant (D = Post - Pre) and tests whether the average difference differs significantly from zero.

### 4.4 One-Way Analysis of Variance (One-Way ANOVA)

When comparing the means of a continuous variable across **three or more independent groups** (e.g., Comparing mean investment volume across three Education groups: High School, Bachelor's, and Master's), running multiple t-tests causes **Type I Error Inflation** (Familywise Error Rate: 1 - (1 - 0.05)^k). One-Way ANOVA tests whether any significant differences exist across all groups simultaneously using a single omnibus **F-statistic**.

**One-Way ANOVA F-Ratio Logic:**
F = Mean Square Between Groups (MS_Between) / Mean Square Within Groups (MS_Within)

• *MS_Between:* Measures variation between group means (treatment effect + error).
• *MS_Within:* Measures random variation within groups (pure individual differences / error).
• If group means differ significantly, F will be substantially greater than 1.00 (p < 0.05).
• **Eta-Squared (Effect Size):** Sum of Squares Between / Total Sum of Squares.

### 4.5 Post-Hoc Multiple Comparisons in ANOVA

A statistically significant ANOVA F-test confirms that at least one group mean differs from another, but it does not specify which exact pairs of groups differ. Researchers execute **Post-Hoc Tests** to conduct pairwise comparisons while strictly controlling familywise error:

- **Equal Variances Assumed (Levene's p > 0.05):**
  - *Tukey's HSD (Honestly Significant Difference):* The premier, balanced post-hoc test when group sample sizes are approximately equal.
  - *Scheffe Test:* Highly conservative; recommended when group sample sizes are unequal or complex comparisons are tested.
  - *Bonferroni Correction:* Adjusts alpha level per test; ideal for a limited number of planned pairwise contrasts.
  - *Duncan's Multiple Range Test:* A stepwise range test that groups means into homogeneous subsets.

DegreeLive

- **Equal Variances Not Assumed (Levene's p ≤ 0.05):**
  - *Games-Howell Test:* Widely considered the most powerful and robust post-hoc test when group variances and sample sizes are unequal.
  - *Tamhane's T2:* A conservative post-hoc alternative under variance inequality.

## 5. Advanced Multivariate Extensions: Two-Way ANOVA, MANOVA, and ANCOVA

To address complex, real-world business questions involving multiple independent factors, confounding covariates, or multiple dependent variables simultaneously, SPSS provides advanced extensions of general linear models:

### 1. Two-Way ANOVA (Factorial Design)

Evaluates the simultaneous effect of two categorical independent variables (Factors A and B) on a continuous dependent variable.
• *Main Effect of Factor A:* Tests whether Factor A significantly influences Y averaged across all levels of Factor B.
• *Main Effect of Factor B:* Tests whether Factor B significantly influences Y averaged across all levels of Factor A.
• *Interaction Effect (A * B):* Tests whether the effect of Factor A depends upon the level of Factor B. In SPSS Profile Plots, intersecting non-parallel lines visually confirm a significant interaction.

### 2. Multivariate Analysis of Variance (MANOVA)

Evaluates group differences across **two or more correlated continuous dependent variables simultaneously**.
• *Advantages:* Protects against Type I error inflation from multiple ANOVAs; detects multivariate patterns invisible to univariate tests.
• *Multivariate Test Statistics:* **Wilks' Lambda** (most common), **Pillai's Trace** (most robust to assumption violations), Hotelling's Trace, and Roy's Largest Root.
• *Assumption:* Tested via **Box's M Test** for equality of covariance matrices.

### 3. Analysis of Covariance (ANCOVA)

Combines ANOVA with linear regression by statistically controlling for the confounding influence of one or more continuous baseline variables (Covariates).
• *Function:* Removes error variance attributed to the covariate, increasing statistical power and generating **Adjusted Marginal Means**.
• *Assumption:* Homogeneity of regression slopes (the relationship between the covariate and dependent variable must be equal across all treatment groups).

### 4. Non-Parametric Equivalents

When parametric assumptions of normality and homogeneity are irrecoverably violated:
• *Mann-Whitney U Test:* Non-parametric equivalent of Independent Samples T-Test.
• *Wilcoxon Signed-Rank Test:* Non-parametric equivalent of Paired Samples T-Test.
• *Kruskal-Wallis H Test:* Non-parametric equivalent of One-Way ANOVA.
• *Chi-Square Test of Independence:* Categorical association.

Degreelive

## 6. Academic Reporting: APA Style Formatting & Guidelines

To adhere to high scholarly publication standards, statistical findings from inferential analyses must be reported in accordance with the American Psychological Association (APA 7th Edition) guidelines:

APA Style Reporting Templates Scholarly Standards

**Reporting Independent Samples T-Test:**
*"An independent-samples t-test was conducted to compare trading volume between male and female investors. Levene's test confirmed equality of variances ($F = 1.42$, $p = 0.235$). Male investors reported significantly higher monthly trading volumes ($M$ = INR 45,200, $SD$ = 12,300) than female investors ($M$ = INR 38,100, $SD$ = 11,500), $t(248) = 4.69$, $p < 0.001$, 95% CI [4,120, 10,080], Cohen's $d = 0.59$."*

**Reporting One-Way ANOVA:**
*"A one-way ANOVA revealed a statistically significant difference in investment returns across the three educational tiers, $F(2, 297) = 8.74$, $p < 0.001$, partial eta-squared = 0.056. Post-hoc comparisons using Tukey's HSD test indicated that participants holding a Master's degree ($M = 14.8\%$, $SD = 3.2\%$) achieved significantly higher returns than those with a High School diploma ($M = 11.2\%$, $SD = 3.8\%$, $p < 0.001$)."*

**Reporting Multiple Regression:**
*"A multiple linear regression analysis was conducted to predict investment volume based on financial literacy, risk tolerance, and household income. The overall model was statistically significant, $F(3, 246) = 28.65$, $p < 0.001$, explaining 25.9% of the variance in investment volume (R-squared = 0.259, Adjusted R-squared = 0.250). Risk tolerance was the strongest significant predictor (Beta = 0.38, $t = 6.12$, $p < 0.001$), followed by financial literacy (Beta = 0.24, $t = 3.88$, $p < 0.001$)."*

**Reporting Exploratory Factor Analysis (EFA):**
*"Principal Component Analysis with Varimax rotation was performed on the 16 survey items. The Kaiser-Meyer-Olkin measure verified sampling adequacy, KMO = 0.862, and Bartlett's test of sphericity was significant, Chi-Square(120) = 1,482.30, $p < 0.001$. Three components had eigenvalues exceeding Kaiser's criterion of 1.00 and in combination explained 62.4% of the total variance."*

## Summary Synthesis

| Statistical Technique | Core Research Purpose | SPSS Menu Navigation | Key Output Metrics to Report |
|---|---|---|---|
| **Pearson Correlation** | Evaluates linear association between two continuous variables. | `Analyze -> Correlate -> Bivariate` | Pearson r, Sig. (2-tailed), N. |
| **Partial Correlation** | Assesses relationship between two variables while controlling for a third. | `Analyze -> Correlate -> Partial` | Partial r, df, p-value. |
| **Multiple Regression** | Predicts a continuous outcome from multiple independent predictors. | `Analyze -> Regression -> Linear` | R-Square, Adjusted R-Square, F-stat, Beta, t-stat, VIF, Durbin-Watson. |
| **Exploratory Factor Analysis** | Reduces data into underlying latent construct dimensions. | `Analyze -> Dimension Reduction -> Factor` | KMO, Bartlett's Chi-Square, Eigenvalues > 1, |

| Statistical Technique | Core Research Purpose | SPSS Menu Navigation | Key Output Metrics to Report |
|---|---|---|---|
| | | | Scree plot, Rotated Loadings. |
| **Independent T-Test** | Compares continuous means between two unrelated groups. | `Analyze -> Compare Means -> Independent-Samples T Test` | Levene's F & Sig., t-value, df, Mean Difference, 95% CI. |
| **Paired Samples T-Test** | Compares repeated measures from the same respondents. | `Analyze -> Compare Means -> Paired-Samples T Test` | Mean paired difference, SD, t-value, df, p-value. |
| **One-Way ANOVA** | Compares continuous means across three or more groups. | `Analyze -> Compare Means -> One-Way ANOVA` | Levene's test, Between/Within MS, F-stat, Tukey/Games-Howell post-hoc. |
| **Two-Way ANOVA** | Evaluates main effects and interaction between two categorical factors. | `Analyze -> General Linear Model -> Univariate` | F-statistic for Factor A, Factor B, and Interaction (A*B), Profile Plots. |
| **MANOVA** | Tests group differences across multiple continuous DVs simultaneously. | `Analyze -> General Linear Model -> Multivariate` | Wilks' Lambda, Pillai's Trace, Box's M test, Multivariate F. |
| **ANCOVA** | Compares group means while statistically controlling continuous covariates. | `Analyze -> General Linear Model -> Univariate` | Covariate F-stat, Adjusted Group Means, Homogeneity of Slopes. |

Degr

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

DegreeLive