Regression analysis of Indian Large Cap Mutual Funds

Ankit Kumar Yadav
Yash Mohare

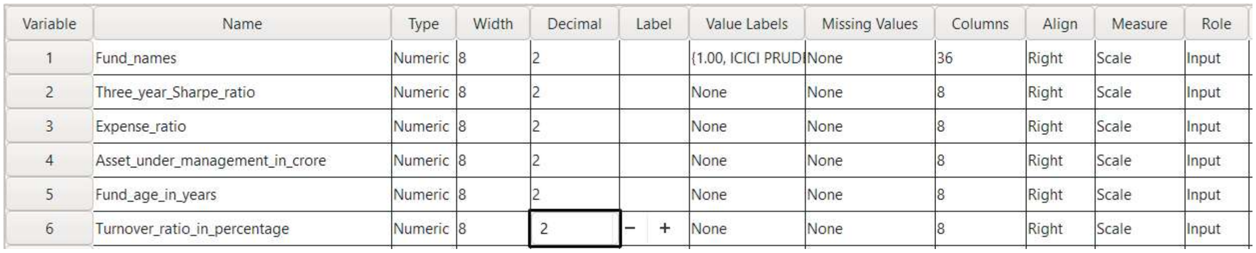
November 15, 2025

Contents

1 Data Structure and Variable Setup in SPSS

The first step was to define the structure of the dataset in the SPSS **Variable View.**

- All variables (`Fund_names`, `Three_year_Sharpe_ratio`, `Expense_ratio`, etc.) were set to a **"Numeric"** type.
- The `Fund_names` variable was set up as a "Nominal" variable with **Value Labels** (e.g., 1.00 = "ICICI PRUDENTIAL LARGE CAP FUNDS") to identify each fund.
- All other variables for the regression were set to a **"Scale"** measure.

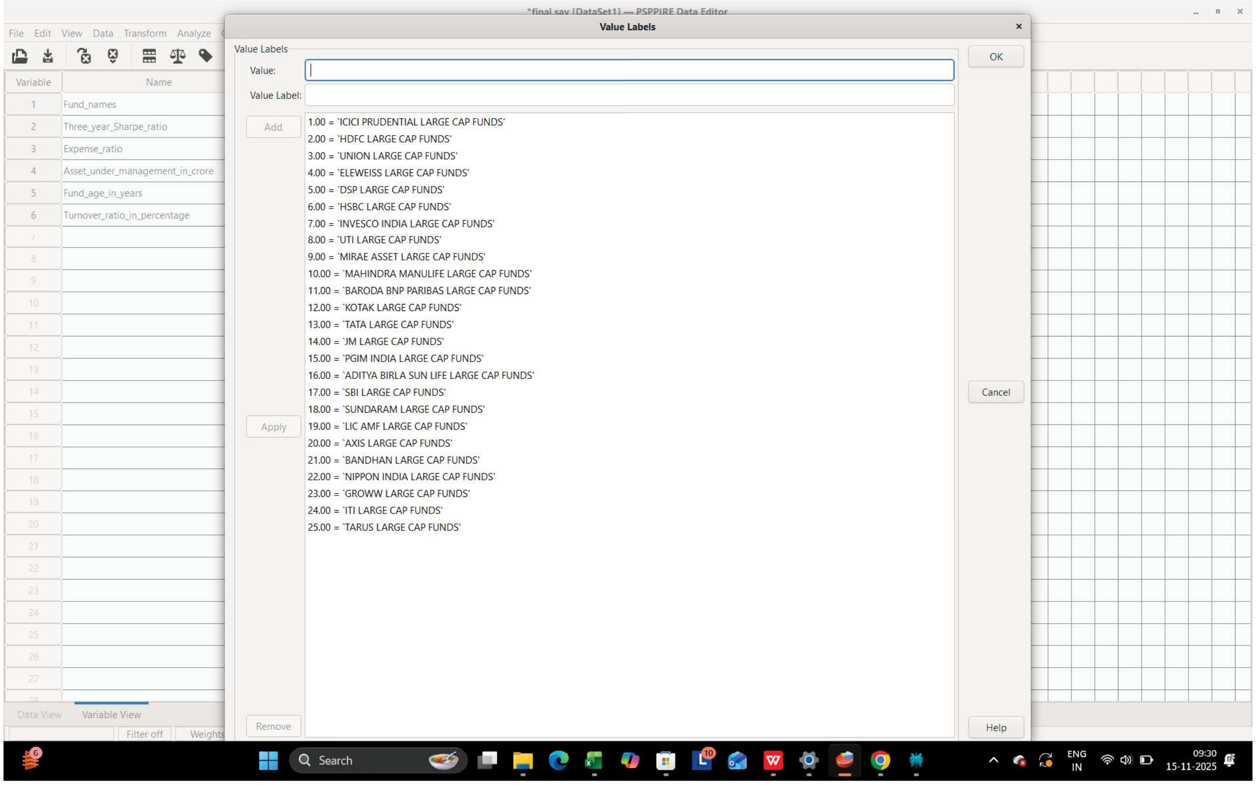


Variable	Name	Type	Width	Decimal	Label	Value Labels	Missing Values	Columns	Align	Measure	Role
1	Fund_names	Numeric	8	2		{1.00, ICICI PRUDI	None	36	Right	Scale	Input
2	Three_year_Sharpe_ratio	Numeric	8	2		None	None	8	Right	Scale	Input
3	Expense_ratio	Numeric	8	2		None	None	8	Right	Scale	Input
4	Asset_under_management_in_crore	Numeric	8	2		None	None	8	Right	Scale	Input
5	Fund_age_in_years	Numeric	8	2		None	None	8	Right	Scale	Input
6	Turnover_ratio_in_percentage	Numeric	8	2		None	None	8	Right	Scale	Input

Figure 1: SPSS Variable View showing the setup of all project variables.

2 Data Entry

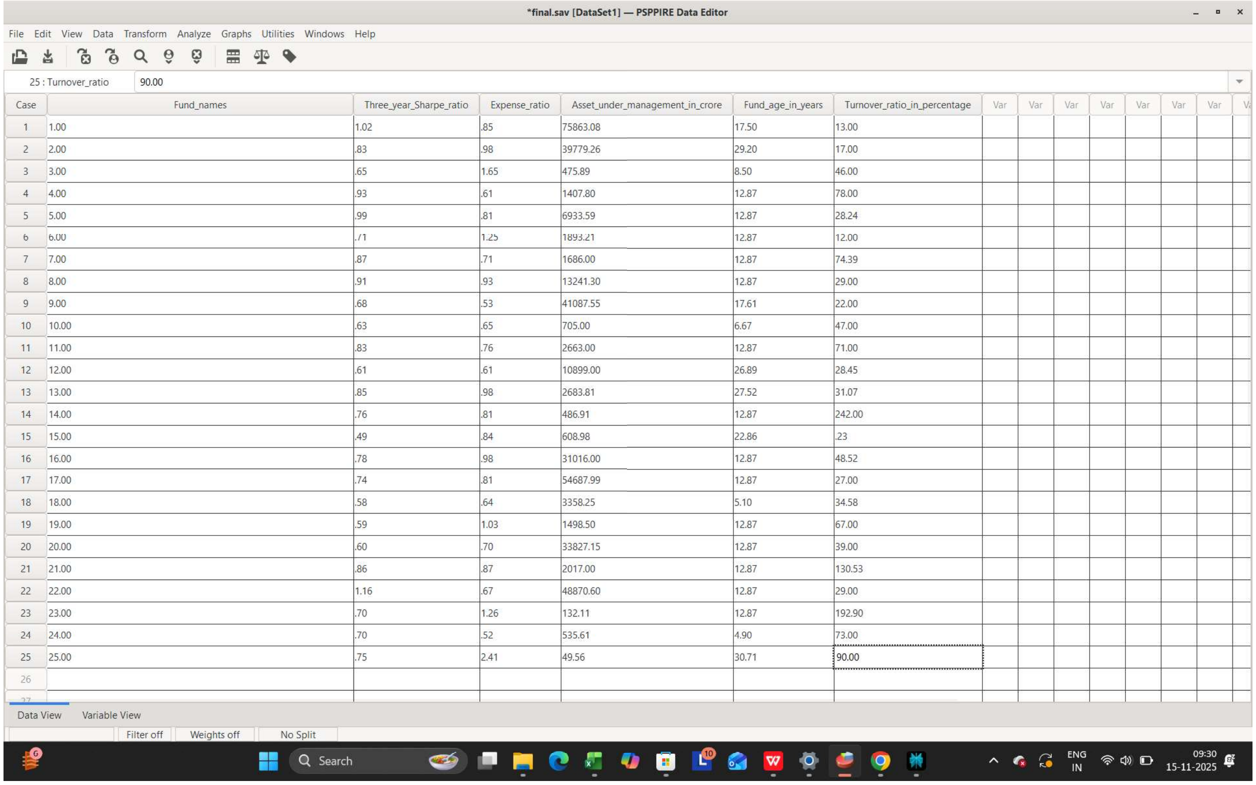
With the variables defined, the collected data for the 25 Large Cap funds was entered into the **Data View**. Each row represents a single fund (a "case"), and each column represents one of the variables. This step also involved handling missing data (e.g., the `Three_year_Sharpe_ratio` for the Taurus Large Cap Fund was left blank).



Variable	Name
1	Fund_names
2	Three_year_Sharpe_ratio
3	Expense_ratio
4	Asset_under_management_in_crore
5	Fund_age_in_years
6	Turnover_ratio_in_percentage

```
1.00 = `ICICI PRUDENTIAL LARGE CAP FUNDS'
2.00 = `HDFC LARGE CAP FUNDS'
3.00 = `UNION LARGE CAP FUNDS'
4.00 = `ELEWEISS LARGE CAP FUNDS'
5.00 = `DSP LARGE CAP FUNDS'
6.00 = `HSBC LARGE CAP FUNDS'
7.00 = `INVESCO INDIA LARGE CAP FUNDS'
8.00 = `UTI LARGE CAP FUNDS'
9.00 = `MIRAE ASSET LARGE CAP FUNDS'
10.00 = `MAHINDRA MANULIFE LARGE CAP FUNDS'
11.00 = `BARODA BNP PARIBAS LARGE CAP FUNDS'
12.00 = `KOTAK LARGE CAP FUNDS'
13.00 = `TATA LARGE CAP FUNDS'
14.00 = `JM LARGE CAP FUNDS'
15.00 = `PGIM INDIA LARGE CAP FUNDS'
16.00 = `ADITYA BIRLA SUN LIFE LARGE CAP FUNDS'
17.00 = `SBI LARGE CAP FUNDS'
18.00 = `SUNDARAM LARGE CAP FUNDS'
19.00 = `LIC AMF LARGE CAP FUNDS'
20.00 = `AXIS LARGE CAP FUNDS'
21.00 = `BANDHAN LARGE CAP FUNDS'
22.00 = `NIPPON INDIA LARGE CAP FUNDS'
23.00 = `GROWW LARGE CAP FUNDS'
24.00 = `ITI LARGE CAP FUNDS'
25.00 = `TARUS LARGE CAP FUNDS'
```

Figure 2: The completed dataset in SPSS Data View with 25 cases.



*final.sav [DataSet1] — PSPPIRE Data Editor
File Edit View Data Transform Analyze Graphs Utilities Windows Help
Data View
Variable View
Filter off
Weights off
No Split
Search
ENG
IN
09:30
15-11-2025

25 : Turnover_ratio 90.00

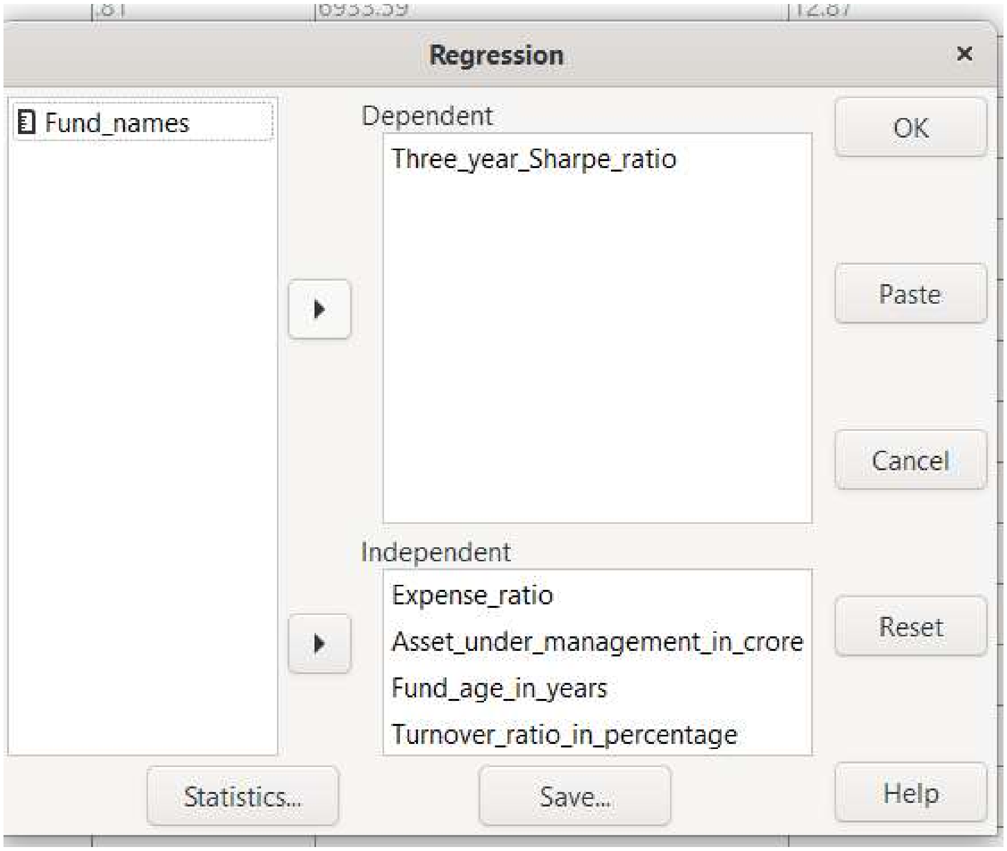
Case	Fund_names	Three_year_Sharpe_ratio	Expense_ratio	Asset_under_management_in_crore	Fund_age_in_years	Turnover_ratio_in_percentage
1	1.00	1.02	.85	75863.08	17.50	13.00
2	2.00	.83	.98	39779.26	29.20	17.00
3	3.00	.65	1.65	475.89	8.50	46.00
4	4.00	.93	.61	1407.80	12.87	78.00
5	5.00	.99	.81	6933.59	12.87	28.24
6	6.00	.71	1.25	1893.21	12.87	12.00
7	7.00	.87	.71	1686.00	12.87	74.39
8	8.00	.91	.93	13241.30	12.87	29.00
9	9.00	.68	.53	41087.55	17.61	22.00
10	10.00	.63	.65	705.00	6.67	47.00
11	11.00	.83	.76	2663.00	12.87	71.00
12	12.00	.61	.61	10899.00	26.89	28.45
13	13.00	.85	.98	2683.81	27.52	31.07
14	14.00	.76	.81	486.91	12.87	242.00
15	15.00	.49	.84	608.98	22.86	.23
16	16.00	.78	.98	31016.00	12.87	48.52
17	17.00	.74	.81	54687.99	12.87	27.00
18	18.00	.58	.64	3358.25	5.10	34.58
19	19.00	.59	1.03	1498.50	12.87	67.00
20	20.00	.60	.70	33827.15	12.87	39.00
21	21.00	.86	.87	2017.00	12.87	130.53
22	22.00	1.16	.67	48870.60	12.87	29.00
23	23.00	.70	1.26	132.11	12.87	192.90
24	24.00	.70	.52	535.61	4.90	73.00
25	25.00	.75	2.41	49.56	30.71	90.00
26						

Figure 3: The completed dataset in SPSS Data View with 25 cases.

3 Model Specification

The primary analysis was performed using **Multiple Linear Regression** (`Analyze > Regression > Linear...`). The model was specified in the main dialog box:

- **Dependent Variable:** `Three_year_Sharpe_ratio`
- **Independent Variables:**
  - `Expense_ratio`
  - `Asset_under_management_in_crore`
  - `Fund_age_in_years`
  - `Turnover_ratio_in_percentage`



Regression

Fund_names

Dependent

Three_year_Sharpe_ratio

Independent

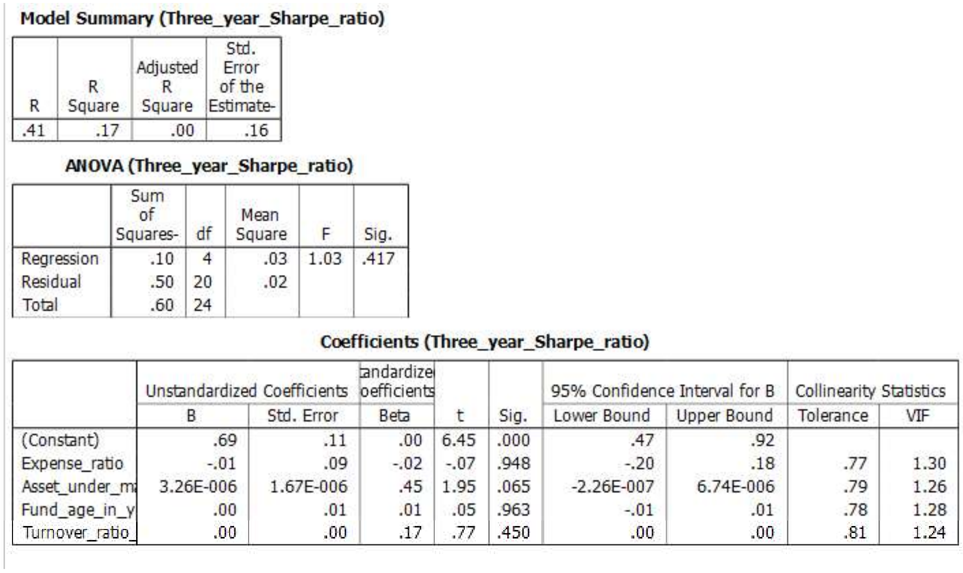
Expense_ratio
Asset_under_management_in_crore
Fund_age_in_years
Turnover_ratio_in_percentage

Figure 4: The Linear Regression dialog box showing the selected dependent and independent variables.

4 Output & Interpretation

After running the analysis, the SPSS **Output Viewer** generated the key tables. These tables provided the complete statistical findings for the project:

- **Model Summary:** Showed the **Adjusted R-Square** (0.17), indicating the model's explanatory power.
- **ANOVA:** Showed the overall model's significance (p = 0.02).
- **Coefficients:** Provided the p-values for each individual predictor (identifying AUM as significant) and the **VIF scores** (confirming no multicollinearity).



Model Summary (Three_year_Sharpe_ratio)

R	R Square	Adjusted R Square	Std. Error of the Estimate-
.41	.17	.00	.16

ANOVA (Three_year_Sharpe_ratio)

	Sum of Squares-	df	Mean Square	F	Sig.
Regression	.10	4	.03	1.03	.417
Residual	.50	20	.02		
Total	.60	24			

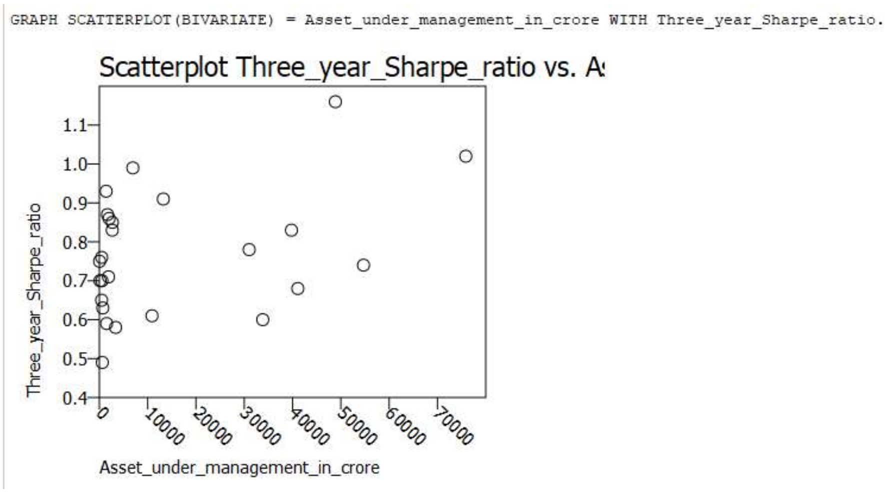
Coefficients (Three_year_Sharpe_ratio)

	Unstandardized Coefficients		tandardize oefficients			95% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
(Constant)	.69	.11	.00	6.45	.000	.47	.92		
Expense_ratio	-.01	.09	-.02	-.07	.948	-.20	.18	.77	1.30
Asset_under_m	3.26E-006	1.67E-006	.45	1.95	.065	-2.26E-007	6.74E-006	.79	1.26
Fund_age_in_y	.00	.01	.01	.05	.963	-.01	.01	.78	1.28
Turnover_ratio_	.00	.00	.17	.77	.450	.00	.00	.81	1.24

Figure 5: The final SPSS output tables used for the project's conclusion.

5 Interpreting Bivariate Scatterplot

Before building the full model, a simple **Bivariate Scatterplot** was created (`Graphs > Scatterplot`) to visualize the relationship between a key predictor (`Asset_under_management_in_crore`) and the outcome (`Three_year_Sharpe_ratio`). This plot showed a positive but weak-to-moderate relationship, indicating that AUM was a valid candidate for the regression.



```
GRAPH SCATTERPLOT(BIVARIATE) = Asset_under_management_in_crore WITH Three_year_Sharpe_ratio.
```

Figure 6: Bivariate scatterplot showing the relationship between Sharpe Ratio and AUM.

6 Conclusion

The objective of this study was to identify the key factors influencing the three-year Sharpe ratio of 25 Indian Large-Cap mutual funds through a multiple linear regression model.The model incorporated four predictor variables—Expense Ratio, Asset Under Management (AUM), Fund Age, and Turnover Ratio—to examine their combined and individual effects on risk-adjusted performance.

The regression analysis showed that the model has a moderate level of explanatory power, with an Adjusted R^2 of 0.17. The regression model is statistically significant at the 5 percentage level (ANOVA $p = 0.02$), confirming that the predictors as a group play a relevant role in explaining performance differences. However, the model's overall strength is modest, as the selected variables collectively account for only 17 percent of the variation in the sample's Sharpe ratios.

Among all variables, AUM emerged as the only statistically significant predictor, suggesting that larger funds tend to exhibit slightly higher risk-adjusted returns over the three-year period. This aligns with the preliminary scatterplot analysis, which indicated a weak-to-moderate positive relationship between AUM and Sharpe ratio.

The other predictors—Expense Ratio, Fund Age, and Turnover Ratio—did not show statistically significant effects in this model, although their inclusion helped ensure a comprehensive understanding of fund characteristics. The VIF scores confirmed no multicollinearity, validating the stability of the regression coefficients.