Analysis Report

This report is structured as follows.

Contents

R Script Summary

The script performs an extensive data analysis workflow involving multiple steps:

1. Library Importation: Several libraries such as openxlsx, dplyr, tidyr, readr, and readxl are imported for data manipulation, transformation, and I/O operations.
2. Data Import: Data from an Excel file 00_Dataset_Regression_Final.xlsx is read into a dataframe df.
3. Data Transformation: The NewProductIntroduction column is converted to numeric.
4. Aggregation: Data is grouped by Name_Compustat and aggregated to calculate means for specific columns.
5. Descriptive Statistics: The script calculates means and standard deviations for the aggregated columns.
6. Data Reshaping: Summary statistics are reshaped into long and wide formats for easier analysis.
7. Frequency Tables: Crosstabulations for variables like Liberalism_Label, SIC, and Gender are generated, including frequency and percentage distributions.
8. Outlier Analysis: Z-score-based outlier identification is performed for several variables.
9. Normality Tests: Skewness, kurtosis, and Shapiro-Wilk tests are calculated to assess the normality of the distribution of variables.
10. Data Transformation: Log transformations are performed for selected variables, and sparse SIC categories are labeled as 'Other'.
11. Modeling: Linear mixed-effects models are fitted for variables AD_Intensity and NewProductIntroduction using the lme4 package, with further p-value adjustments made using the afex package.
12. Exporting Results: Summary statistics and model results are written to an Excel file named Results.xlsx.

Descritive Statistics

The descriptive statistics table presents the means and standard deviations for five key variables. The average Liberalism Index was found to be 0.592 with a standard deviation of 0.427. Mean AD Intensity exhibited a mean of 0.056 and a standard deviation of 0.101. Mean New Product Introduction had a mean of 9.704 and a considerable standard deviation of 19.638, indicating a wide spread of the data. Mean Total Assets and Mean ROA showed means of 26018.154 and 0.032, with standard deviations of 41548.331 and 0.113, respectively.

Variable	mean	sd
mean_Liberalism_Index	0.592	0.427
mean_AD_Intensity	0.056	0.101
mean_NewProductIntroduction	9.704	19.638
mean_Total_Assets	26018.154	41548.331
mean_ROA	0.032	0.113

A frequency distribution of the Liberalism Label showed that the majority of the sample could be classified as 'Fully liberal' (50.746%), followed by 'Fully conservative' (35.821%). The remaining categories, namely 'Moderate,' 'Slightly conservative,' 'Slightly liberal,' and 'Strongly liberal,' had relatively fewer occurrences, making up less than 10% of the sample each.

Liberalism_Label	Freq	Percentage
Fully conservative	24	35.821
Fully liberal	34	50.746
Moderate	3	4.478
Slightly conservative	1	1.493
Slightly liberal	4	5.970
Strongly liberal	1	1.493

Normality Stats

Assessment of data normality showed that all variables, except Liberalism Index, had significant skewness and kurtosis., which led them to be substantially deviant from normality, as indicated by the Shapiro Wilk's tests (columns W and P_value) Consequently, these variables were log-transformed to improve normality.

Variable	Skewness	Kurtosis	W	P_Value
Liberalism_Index	-0.467	1.317	0.703	0.000
AD_Intensity	4.528	24.831	0.412	0.000
NewProductIntroduction	2.940	11.736	0.546	0.000
Total_Assets	2.998	12.684	0.597	0.000
ROA	-5.713	52.075	0.557	0.000

Linear Mixed Model – Ad Intensity

In the linear mixed model analysis focused on AD Intensity, most variables did not show a significant effect, as indicated by the p-values above 0.05. Notably, Total Assets had a significant effect with a p-value of 0.008. It's worth mentioning that several sectors, represented by SIC columns, had to be dropped due to low sample sizes and the presence of missing data in other variables, which led to rank-deficient model matrices. Nonetheless, the effects of some sectors are still controlled in the model.

Variable	Coefficient	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Liberalism_Index	0.040	0.000	0.000	1	26.441	1.121	0.299
RD_Intensity	-0.049	0.000	0.000	1	112.200	0.174	0.677
Total_Assets	-0.015	0.002	0.002	1	54.920	7.522	0.008
ROA	0.038	0.001	0.001	1	90.641	3.565	0.062
SIC_SIC48	-0.033	0.000	0.000	1	27.063	0.152	0.700
SIC_SIC73	0.003	0.000	0.000	1	36.552	0.004	0.953

Linear Mixed Model – New Product Introduction

For the New Product Introduction dependent variable, Total Assets again showed a significant effect with a p-value of 0.001. Other variables such as Liberalism Index and RD Intensity did not reach statistical significance. If a 0.1 significance level is considered, one could state that Liberalism Index has a significant effect on New Product Introduction ($p < 0.1$). The positive coefficient of 0.634 indicates the higher the liberalism, the higher the number of new product introductions.

Similar to the AD Intensity model, some sectors had to be omitted due to low sample sizes and missing data, but their effects were partially captured in the analysis.

Variable	Coefficient	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Liberalism_Index	0.634	0.962	0.962	1	35.470	2.879	0.099
RD_Intensity	1.045	0.376	0.376	1	142.490	1.124	0.291
Total_Assets	0.252	3.920	3.920	1	37.647	11.732	0.001
ROA	-0.298	0.079	0.079	1	143.587	0.235	0.628
SIC_SIC48	-0.114	0.004	0.004	1	37.836	0.013	0.910
SIC_SIC73	0.923	0.817	0.817	1	40.982	2.445	0.126