Author: Jiwon Jeong, PhD in Demography

Data Analytics Portfolio

Major Analytical Datasets	- National-level high-dimensional dataset: approximately 27,300 rows × 4,000 columns (≈100 million cells) - National-level high-dimensional dataset: approximately 25,000 rows × 1,100 columns (≈27.5 million cells) - Multi-country medium-scale dataset (self-constructed): approximately 1,000 rows × 320 columns (≈320,000 cells)
Languages	- R, Python, SQL

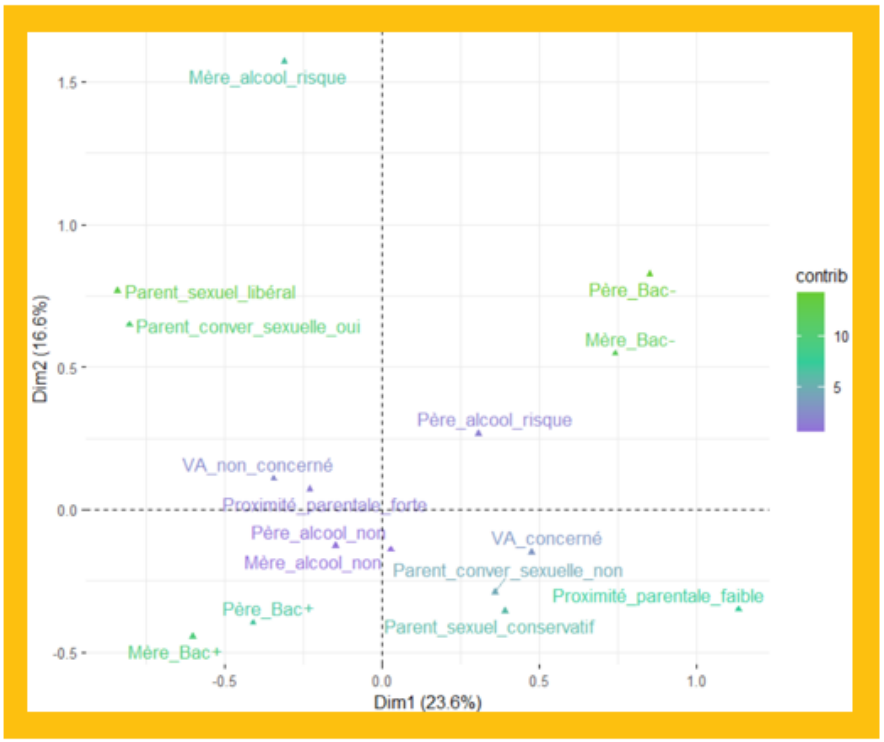
 # The analytical cases in the portfolio follow the logical framework of applied data analysis (data exploration → variable identification and classification → correlation analysis → multiple regression analysis)

Case 1) Identifying Multivariate Patterns Using Multiple Correspondence Analysis (MCA)

- Identified associations and distributions among various categorical variables within the biplot dimensions
- Determined statistical significance of variable groups positioned closely within each dimension

- Application areas:
  - Establishing segment-specific strategies by clustering consumer behavior patterns or consumer characteristics related to purchasing behavior
  - Identifying combinations of multivariate factors associated with cost reduction
  - Exploring complex combinations among diverse performance indicators that influence ROI and QCD (Quality, Cost, Delivery)

(Source : PhD dissertation)

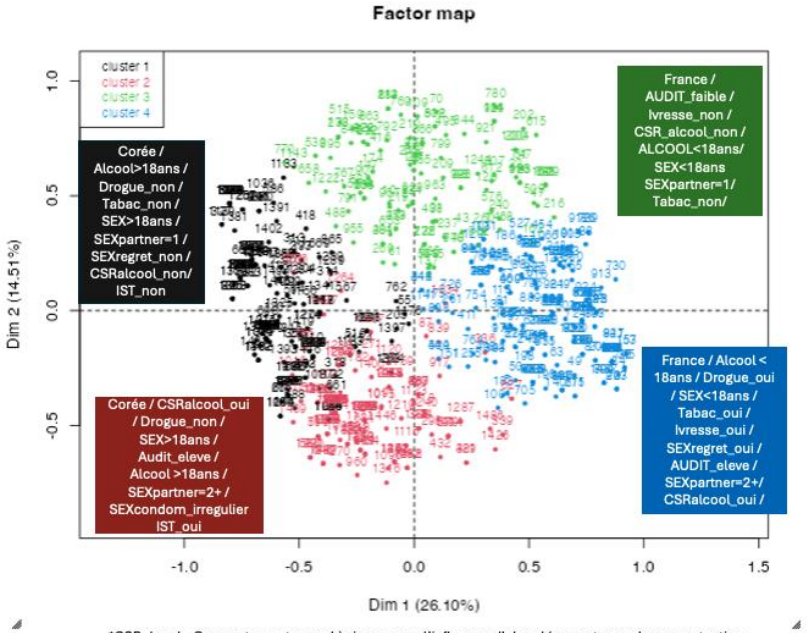


Case 2) Profiling via Hierarchical Clustering Analysis

- Clustered observations that share similar characteristics within the biplot dimensions
- Derived profiling based on variables with statistically significant contributions within each cluster

- Application areas:
  - Performing multivariate segmentation
  - Developing differentiated marketing strategies or KPIs tailored to each segment

(Source : Research proposal submitted to French National Center for Scientific Research, CNRS)



**CSRalcool – Comportement sexuel à risque sous l'influence d'alcool (rapport sexuel sans protection, sexe occasionnel, sexe transactionnel)*

Case 3) Partial Correlation Analysis to Examine Linear Relationships While Controlling for Influencing Variables

- Measured the linear relationship between two quantitative variables while limiting the influence of control variables
- Application areas:
  - Evaluating the correlation between investment and revenue growth while controlling for multiple influencing factors (i.e., assessing the true association between investment and revenue)
  - Selecting performance indicators with strong statistical correlations prior to building a multiple regression model during ROI evaluation tool development

(Source : International conference proceedings and poster)

Spearman Correlation
-1.0 0.0 1.0

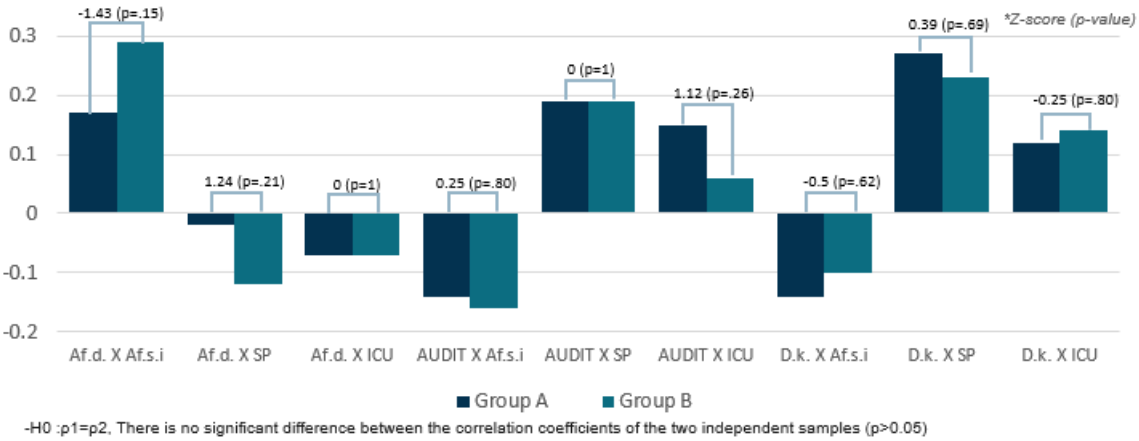
	Age of A.I.	AUDIT-C score	Number of S.D.		Age of A.I.	AUDIT-C score	Number of S.D.
Age at FSX	0.17***	-0.14***	-0.14***	Age at FSX	0.28***	-0.16***	-0.10***
Number of S.P.	-0.02	0.19***	0.26***	Number of S.P.	-0.12***	0.19***	0.23***
Inconsistent CU	-0.07***	0.15***	0.12***	Inconsistent CU	-0.07**	0.06*	0.14***

-p*<0.05, p**<0.01, p***<0.001, Control variables: age, gender, occupation, level of education, subjective socioeconomic status

Case 4) Comparing Correlation Coefficients Between Two Independent Groups (z-test)

- Compared standardized z-scores for partial correlation coefficients between two groups
- Application areas:
  - Assessing whether the impact of a new system on defect-rate reduction is consistent across all factories
  - Determining whether a standardized solution should be established for system implementation

(Source : International conference proceedings and poster)



-H0 :p1=p2, There is no significant difference between the correlation coefficients of the two independent samples (p>0.05)

Case 5) Developing Multiple Regression Models with Multivariate Predictors and Interaction Terms

- Performed multiple regression analysis to examine relationships between various predictors and the outcome variable while controlling for confounding variables
- Generated predicted values for both numerical and categorical outcome variables; incorporated interaction terms to derive category-specific predictions
- Application areas:
  - Providing quantitative predictions of purchase likelihood, revisit probability, and other outcomes based on diverse digital behavioral data within D2C platforms
  - Estimating the increase in sales attributable to higher investment levels
  - Measuring the effect of investment on sales growth according to QCD categorical performance indicators (e.g., quality good vs. defective)

(Source : PhD dissertation)