Turas Categorical Key Driver (CatDriver)

Transparent. Trustworthy. Tangible insights.

Key driver analysis for categorical outcomes—what influences tier choice, segment membership, or yes/no decisions. When your outcome is categorical (Satisfied/Neutral/Dissatisfied, Buy/Don't Buy, Promoter/Passive/Detractor), CatDriver uses regression models designed specifically for categorical data.

What This Module Does

Standard regression fails for categorical outcomes because it can predict impossible values. CatDriver uses ordinal and multinomial logistic regression to model probabilities correctly—respecting category boundaries and handling ordered vs. unordered categories appropriately.

The module identifies what drives customers to choose "Very Satisfied" vs. "Somewhat Satisfied," what factors predict purchase (Buy vs. Don't Buy), and what differentiates Promoters from Detractors—with proper statistical handling of categorical data.

What You Get

Statistical Outputs

- **SHAP-based importance scores:** Driver importance adapted for categorical models
- **Regression coefficients:** With confidence intervals for each driver
- **Odds ratios:** How much each driver changes the odds of being in a higher category
- **Individual predictions:** What drives each respondent's category membership

Model Diagnostics

- **McFadden's R^2:** Model fit measure for logistic regression
- **Proportional odds test:** Validates whether ordinal model assumptions are met
- **AIC/BIC:** Model selection criteria

How It Works

Model Selection

2 categories → Binary logistic regression (glm). 3+ ordered categories → Ordinal logistic regression (clm/polr). 3+ unordered categories → Multinomial logistic regression (multinom).

Ordinal Logistic Regression

Models cumulative probabilities: P(Y ≤ Dissatisfied), P(Y ≤ Neutral), P(Y ≤ Satisfied). Uses cumulative link model to respect the ordering of categories. Provides single coefficient per driver (proportional odds assumption).

Odds Ratio Interpretation

OR = 2.51 means: A 1-point increase in that driver more than doubles the odds of being in a higher category. This is multiplicative—easy to communicate to stakeholders.

Packages Used

All packages are peer-reviewed, open-source R packages available on CRAN.

Package	Why We Use It
ordinal::clm	Cumulative link models for ordinal regression—preferred approach
MASS::polr	Fallback ordinal logistic (proportional odds)—R Core Team maintained
nnet::multinom	Multinomial logistic for unordered categories
brglm2	Firth correction for separation issues in sparse data

Strengths

- ✓ **Purpose-built for categorical:** Handles satisfaction scales, purchase decisions, NPS categories correctly
- ✓ **Coefficient interpretation:** Clear driver effects with odds ratios and confidence intervals
- ✓ **Robust to problems:** Bias reduction methods prevent model failures from separation
- ✓ **Automatic model selection:** Chooses binary/ordinal/multinomial based on your data
- ✓ **Individual-level insights:** Shows what drives each person, not just population averages
- ✓ **Canonical implementation:** Uses Venables & Ripley's MASS—R Core Team maintained

Limitations

- ⚠ **More complex:** Requires more statistical knowledge to interpret fully
- ⚠ **Sample size requirements:** Need n≥200 for stable ordinal estimates, n≥300 for multinomial
- ⚠ **Proportional odds assumption:** Ordinal models assume driver effects are constant across thresholds
- ⚠ **Computation time:** SHAP calculation can be slow for very large datasets

Sample Size Requirements

Ordinal logistic: minimum n≥200, recommended n≥300. Multinomial logistic: minimum n≥300, recommended n≥500 (estimates K-1 sets of coefficients). Rule of thumb: 10-20 observations per predictor variable.

Best Use Cases

Ideal For

- ✓ NPS driver analysis (Promoter/Passive/Detractor)
- ✓ Purchase intent drivers (Will buy/Might/Won't)
- ✓ Satisfaction category drivers (Very/Somewhat/Not Satisfied)
- ✓ Segment classification (what makes someone fit a segment)

Not Ideal For

- ⚠ Continuous outcomes (use KeyDriver instead)
- ⚠ Very small samples (<150 respondents)
- ⚠ When simple correlation analysis suffices
- ⚠ Real-time applications (SHAP calculation can be slow)

Key Takeaways

- ✓ Use CatDriver when your outcome has distinct categories—not a continuous scale
- ✓ Odds ratios (OR) show how much each driver changes the odds of being in a higher category
- ✓ Proportional odds test validates whether the ordinal model is appropriate
- ✓ Firth correction prevents model failures when some categories have few observations

The Bottom Line

CatDriver brings rigorous categorical regression to market research. When your outcome is categorical and you need to understand what drives it, this module provides sophisticated yet interpretable analysis. Using ordinal and multinomial logistic regression with proper handling of survey data, CatDriver tells you exactly which drivers move people between categories—with full statistical validation.

About The Research LampPost

Turas is developed and delivered by **The Research LampPost**, an organisational member of the Southern African Marketing Research Association (SAMRA). Duncan Brett is a SAMRA accredited researcher with 30+ years of market research experience. If you have questions about whether Turas is right for your project, if there is interest in the statistical documentation or module-specific information, or you would like a copy of our credentials, please get in touch at duncan@researchlamppost.co.za.

Ready to discuss your project?

duncan@researchlamppost.co.za