

# Turas

*Transparent. Trustworthy. Tangible insights.*

Technical Specification & Statistical Documentation



**Platform:** R 4.0+ with renv for full reproducibility | **Modules:** 11 production-ready | **Reliability:** TRS v1.0

*Every calculation uses peer-reviewed, open-source R packages. No proprietary black boxes. All methods auditable.*

## Design Philosophy

**Transparency:** All methods use published R packages with peer-reviewed implementations.

**Defensibility:** Industry-standard methods with documented assumptions and limitations.

**Reproducibility:** renv package management locks exact versions. Your analysis will reproduce identically in five years.

**Data Security:** Your data runs on our machines—never uploaded to AI platforms for processing.

## Turas Reliability Framework (TRS)

Every Turas analysis produces a definitive outcome. No silent failures. No ambiguous results.

Component	Purpose
Guard Layer	Pre-flight validation before any analysis runs
Structured Refusals	Clear error messages with actionable fix guidance
Run State Tracking	Four definitive outcomes: PASS, PARTIAL, REFUSE, ERROR
Atomic File Writes	Prevents corrupt or partial output files

## Module Overview

Module	Primary Packages	Why This Package
⚖️ Weighting	survey	Gold standard—powers US Census, CDC, WHO
📊 Tabs	Base R stats	R Foundation maintained, universally validated
🎯 Confidence	Base R, future	Core stats + parallel processing for bootstrap
🔑 KeyDriver	xgboost, shapviz	Industry ML standard + Lundberg's TreeSHAP
🔒 CatDriver	MASS, nnet	R Core Team maintained, canonical implementations
📈 Tracker	Base R, future	Standard parametric inference, parallelised
⚡️ Conjoint	mlogit	Nobel-cited methodology (McFadden 1974)
↑ MaxDiff	survival, cmdstanr	Mayo Clinic clogit + Gelman's Stan team
🧩 Segment	cluster, poLCA	R recommended packages, proven algorithms
💰 Pricing	pricesensitivitymeter	Purpose-built for Van Westendorp/Gabor-Granger
🔧 AlchemerParser	Base R	Workflow automation, no external dependencies

## Module Specifications

### Weighting

Corrects sample imbalances so your data reflects the population you're studying.

**Package:** *survey* (Thomas Lumley) — The definitive R implementation for complex survey analysis, used by US Census Bureau, CDC, and WHO.

**Methods:** Design weights (cell weighting), rim weights via `calibrate()` with raking/linear/logit calibration, weight trimming, efficiency diagnostics ( $n_{eff}/n$  ratio).

**Foundation:** Deming & Stephan (1940), Kish (1965).

### Tabs

Cross-tabulations with significance testing—the foundation of survey reporting.

**Package:** *Base R stats* — R Foundation maintained, the most validated statistical functions available.

**Methods:** Weighted z-tests for proportions, weighted t-tests for means, chi-square tests, Fisher's exact (small cells), effective sample size (Kish 1965), multiple comparison correction (Bonferroni, Holm).

**Output:** Cross-tabulation matrices, significance markers with p-values, Wilson score CIs, sample sizes (weighted/unweighted).

### Confidence

Quantifies how reliable your estimates are—margin of error, credibility intervals, sample quality.

**Packages:** *Base R, future* — Core statistical functions plus parallel processing for computationally intensive bootstrap methods.

**Proportions:** Normal approximation (MOE), Wilson score, bootstrap percentile (BCa), Bayesian credible intervals.

**Means & NPS:** t-distribution CI, bootstrap percentile, Bayesian credible intervals.

**Sample quality:** Kish effective sample size, DEFF, CI adjustment for design effects, weighted bootstrap.

**Foundations:** Brown, Cai & DasGupta (2001), Kish (1965), Efron & Tibshirani (1993).

### KeyDriver

Identifies which factors most influence your outcome metric (NPS, satisfaction, likelihood to recommend).

**Packages:** *xgboost, shapviz* — Industry-leading gradient boosting (used by Kaggle winners) plus Lundberg's TreeSHAP implementation.

**Methods:** Partial R<sup>2</sup> decomposition (Lindeman, Merenda & Gold 1980), SHAP values for individual-level attribution, importance-performance analysis.

**Advantages:** Handles multicollinearity, captures non-linear effects, provides individual-level explanations..

### CatDriver

Key driver analysis for categorical outcomes—what influences tier choice, segment membership, or yes/no decisions.

**Packages:** *MASS, nnet, brglm2*.

**Methods:** Binary logistic (`glm`), ordinal logistic (`polr`), multinomial logistic (`multinom`), bias-reduced estimation for sparse data.

**Key feature:** Canonical design-matrix mapper ensures correct coefficient-to-level mapping—no string parsing hacks.

**Diagnostics:** Brant test (proportional odds), AIC/BIC, McFadden's R<sup>2</sup>, ROC/AUC. Foundations: Agresti (2002), McCullagh (1980).

### Tracker

Tracks survey metrics across waves—have perceptions or satisfaction metrics actually shifted across waves?

**Packages:** *Base R stats, future* — Standard parametric inference with parallel processing for multi-wave datasets.

**Methods:** Wave-to-wave z-tests and t-tests, trend significance testing, banner trend analysis, question mapping across waves with code changes.

**Quality checks:** Base size stability, demographic composition drift (chi-square), question continuity validation.

## ⚡ Conjoint

Understands preference trade-offs—if you reduce price but remove free shipping, do customers still prefer your offer?

**Package:** *mlogit* (Yves Croissant) — Implements McFadden's (1974) random utility model, the methodology that won the 2000 Nobel Prize in Economics.

**Methods:** Multinomial logit (MNL), conditional logit fallback (survival::clogit).

**Requirements:**  $n \geq 200$  for aggregate, 10-15 choice tasks per respondent.

**Output:** Part-worth utilities (zero-centered), attribute importance, Excel market simulator with sensitivity analysis.

## ⇓ MaxDiff

Reveals true preference rankings—which features, messages, or attributes matter most to your customers.

**Packages:** *survival, cmdstanr* — Mayo Clinic's clogit (30+ years clinical validation) plus Andrew Gelman's Stan team for HB estimation.

**Methods:** Conditional logit for aggregate, Hierarchical Bayes via Stan (HMC sampling) for individual-level utilities.

**Design:** Balanced incomplete block designs via *AlgDesign* with D-optimal efficiency.

**Convergence:** Full diagnostics (Rhat, ESS, divergences). Requirements:  $n \geq 150$  aggregate,  $n \geq 200$  for HB.

**Foundations:** Louviere & Woodworth (1983), Carpenter et al. (2017).

## ✳️ Segment

Discovers natural groups in your customer base—distinct segments that need different approaches.

**Packages:** *cluster, poLCA, MASS* — R's recommended packages for clustering and classification, proven algorithms with decades of validation.

**Methods:** K-means (Hartigan-Wong), Latent Class Analysis for categorical data.

**Optimal K:** Elbow method, silhouette score, Calinski-Harabasz index, BIC for model-based selection.

**Validation:** LDA discrimination, ANOVA/chi-square profiling, Mahalanobis outlier detection. Requirements: min 100 per cluster.

## 💰 Pricing

Determines optimal price points and acceptable ranges—what will customers pay?

**Package:** *pricesensitivitymeter* — Purpose-built for pricing research, implements established methodologies correctly.

**Van Westendorp PSM:** Four-question price perception analysis—OPP, IDP, PMC, PME with Newton-Miller-Smith extension for purchase probability.

**Gabor-Granger:** Sequential purchase intent, demand curves, price elasticity, revenue optimization.

**Foundations:** Van Westendorp (1976), Gabor & Granger (1966), Newton, Miller & Smith (1993).

## Scope & Limitations

Turas is designed for applied market research - solid, defensible analysis.

**Not implemented:** General Bayesian methods (except HB for MaxDiff), causal inference frameworks (IV, RDD), machine learning beyond XGBoost for SHAP, multiple imputation for missing data (listwise deletion is default), real-time/streaming analysis, interactive dashboards.

**Assumption handling:** Assumptions are tested and documented. Normality violations trigger robust alternatives (bootstrap). Outliers flagged with diagnostic plots. Minimum sample sizes enforced—underpowered analyses refused, not run with warnings.

## What You Receive

Every Turas project includes documentation that lets you defend your methodology:

- ✓ **Excel reports** with multiple worksheets, formatted output, significance markers (if requested)
- ✓ **Methodology documentation** — methods used, assumptions tested, limitations acknowledged
- ✓ **R package versions** — exact versions used, enabling reproduction years later
- ✓ **Interpretation guidance** — what the numbers mean in plain language
- ✓ **Technical appendix available** — detailed specifications

## Package Provenance

All packages are CRAN-published with established track records. Combined downloads exceed **400 million**. Key maintainers include:

- **Thomas Lumley** (survey) — University of Auckland
- **Andrew Gelman** (Stan/cmdstanr) — Columbia University
- **Terry Therneau** (survival) — Mayo Clinic
- **Yves Croissant** (mlogit) — Econometrician
- **Venables & Ripley** (MASS) — R Core Team
- **Scott Lundberg** (SHAP methodology)

## The Bottom Line

Turas uses established methods correctly. For applied market research, it provides solid, defensible analysis with full documentation. **No black boxes. No proprietary algorithms.**

## 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 module-specific information, or you would like a copy of our credentials, please get in touch at [duncan@researchlamppost.co.za](mailto:duncan@researchlamppost.co.za).

**Technical inquiries & collaboration**

[duncan@researchlamppost.co.za](mailto:duncan@researchlamppost.co.za)