Package: senseR (via r-universe)

October 8, 2026

Title Proxy Indicator Diagnostic Tool for Analytical and Policy Use

Version 0.1.0

Description Provides statistical diagnostics to evaluate whether proxy indicators reliably represent an unobservable target construct. The main function 'senser()' assesses proxies across multiple dimensions including monotonicity, information content, stability, distributional alignment, and potential bias risk. It prints a concise, interpretable summary suitable for analytical and policy-oriented assessment, without claiming causal inference.

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

Repository https://jokoadenur.r-universe.dev

Date/Publication 2026-02-17 07:32:34 UTC

RemoteUrl https://github.com/jokoadenur/senser

RemoteRef HEAD

RemoteSha 19dbe0ef08f4323f899f1dd6f2b30769898e7830

Contents

senser 2

Index **5**

senser *senser: Proxy Indicator Diagnostic Tool*

Description

`senser()` is a statistical diagnostic function designed to evaluate whether one or more proxy indicators are suitable representations of an underlying construct that cannot be directly observed or measured.

The function assesses each proxy based on multiple statistical dimensions: monotonicity, information content, stability, distributional alignment, bias risk, and dynamic range (sensitivity). The output is a concise, interpretable summary printed directly to the console.

This tool is intended for analytical diagnostics and policy-oriented indicator assessment. It does not claim causal inference.

Usage

```
senser(
  data,
  proxy,
  target,
  lang = c("english", "indonesia"),
  stagnation_cut = 0.01,
  cv_cut = 0.02,
  ceiling_cut = 0.95
)
```

Arguments

`data`	A `data.frame` containing the target construct and proxy variables.
`proxy`	A character vector specifying one or more proxy indicator variable names contained in `data`.
`target`	A character string specifying the target construct variable name contained in `data`.
`lang`	A character string specifying the language of interpretation. Options are `"english"` (default) or `"indonesia"`.
`stagnation_cut`	Numeric scalar. Threshold used to detect stagnation (very small average absolute change). If the average absolute change of the proxy variable is below this value, a structural penalty is applied. Default is 0.01.
`cv_cut`	Numeric scalar. Threshold for the coefficient of variation (CV). If the coefficient of variation of the proxy variable is below this value, a variability penalty is applied. Default is 0.02.
`ceiling_cut`	Numeric scalar. Threshold used to detect ceiling effects. If the ceiling ratio exceeds this value, a ceiling penalty is applied. Default is 0.95.

Details

The stagnation threshold is used to penalize proxy variables that exhibit minimal temporal or cross-sectional variation. Extremely small changes may indicate lack of sensitivity or measurement rigidity.

The coefficient of variation (CV) is defined as the ratio of the standard deviation to the mean. Very low CV values indicate insufficient dispersion, which may reduce the informational usefulness of the proxy variable.

A ceiling effect occurs when observations cluster near the upper bound of the distribution, limiting discriminatory power. If the ceiling ratio exceeds this threshold, the proxy's score is penalized to reflect reduced measurement sensitivity.

The diagnostic score for each proxy is computed using six normalized components:

1. **Monotonicity**: Spearman rank correlation between proxy and target.
2. **Information content**: Proportion of variance explained (R-squared).
3. **Stability**: Sensitivity of regression coefficients across subsamples.
4. **Distributional alignment**: Similarity of standardized distributions using the Kolmogorov–Smirnov statistic.
5. **Bias risk**: Penalization for strong nonlinearity indicating potential proxy distortion.
6. **Dynamic range / Sensitivity**: Penalization if the proxy has very small changes relative to the target (detects ceiling effect or nearly flat proxies).

The final score is calculated as the **median** of all six components, providing a robust measure less sensitive to extreme values.

The score ranges from 0 to 1 and is classified into three categories:

- **Suitable proxy**: score >= 0.70
- **Conditionally suitable**: 0.40 <= score < 0.70
- **Not suitable proxy**: score < 0.40

Interpretation is automatically generated in the selected language.

Value

The function prints a structured diagnostic summary to the console. No object is returned invisibly. This design prioritizes interpretability and ease of use for applied users and policymakers.

Author(s)

Joko Nursiyono (concept)
senseR development team

See Also

`lm`, `cor`, `ks.test`

Examples

```
## Example with multiple proxies
set.seed(123)
df <- data.frame(
  gdp = rnorm(100, 10, 2),
  ntl = rnorm(100, 50, 10),
  road_density = rnorm(100, 3, 0.5),
  mobile_signal = rnorm(100, 70, 15)
)

senser(
  data = df,
  proxy = c("ntl", "road_density", "mobile_signal"),
  target = "gdp",
  lang = "english"
)
```

Index

cor, *3*

ks.test, *3*

lm, *3*

senser, 2