Package: propertycohorts (via r-universe)

November 6, 2026

Type Package

Title Audit Comparability of Property Listing Cohorts

Version 1.0.0

Description Audits two explicitly labelled listing snapshots before comparing their mean values. Reports excluded rows, unsupported cells, common-cohort coverage, and an exact symmetric decomposition into within-cell and composition differences. Produces local HTML and formula-safe CSV reports without runtime dependencies outside base R. These are descriptive audits of supplied observations, not market indices or causal estimates.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 4.1.0)

Imports graphics, utils

URL https://novasia.estate/, https://github.com/novasia-estate/property-cohort-audit

BugReports https://github.com/novasia-estate/property-cohort-audit/issues

Repository https://novasia-estate.r-universe.dev

Date/Publication 2026-11-06 13:09:19 UTC

RemoteUrl https://github.com/novasia-estate/property-cohort-audit

RemoteRef HEAD

RemoteSha e32c148642dfaa632a7e371d66510ca1ebd38a64

Contents

audit_cohorts 2
synthetic_cohorts 3
write_cohort_report 4

Index **5**

audit_cohorts *Audit the Comparability of Two Listing Snapshots*

Description

Reports common-cell coverage and descriptive within-cell and composition components of an observed mean difference.

Usage

```
audit_cohorts(before, after, strata, value_col = "value",
  id_col = "listing_id", unit_col = "unit", min_cell_n = 2L)
## S3 method for class 'cohort_audit'
print(x, ...)
## S3 method for class 'cohort_audit'
plot(x, ...)
```

Arguments

before, after	Data frames representing two supplied snapshots. Each requires a character listing ID, positive finite numeric value, explicit character unit label and the chosen strata columns.
strata	One to six distinct categorical column names. Character or factor values are trimmed and remain case-sensitive. No geographic or semantic normalization is inferred.
value_col, id_col, unit_col	Distinct column names. Numeric-looking text values must be parsed explicitly by the caller. All nonempty unit labels must agree exactly after trimming; no conversion is attempted.
min_cell_n	Positive integer count required in both periods for a cell to be eligible. This threshold does not establish statistical reliability or privacy.
x	A result returned by audit_cohorts.
...	For plotting, additional arguments to graphics::barplot. Otherwise unused.

Details

Duplicate IDs are checked separately within each snapshot. All occurrences of a duplicate are excluded. The same ID may legitimately appear across periods, but this is a repeated-cross-section comparison, not a paired-property estimator. Exclusions contain original row numbers and one primary reason per row.

Eligible cells meet the minimum count in both snapshots. Cell weights are separately normalized eligible observation counts per period. For each eligible cell with means m0 and m1 and weights w0 and w1, the within component is (w0+w1)/2 times (m1-m0), and the composition component is (m0+m1)/2 times (w1-w0). Sums of these components equal the common-cell mean difference. These are descriptive arithmetic components, not causal effects.

Raw means use all valid observations, while common means use only eligible common cells. When coverage is incomplete the common difference need not equal the raw difference. Unsupported cells remain visible and are not extrapolated. Empty or unsupported samples yield unavailable metrics rather than zero effects.

Unit labels are caller-supplied metadata, not verified currency or measurement semantics. Verify valuation basis, dates, currencies and area definitions before analysis. Asking prices do not establish achieved sale prices. This audit is not a market index, valuation or probability-sample inference.

Value

An object of class `cohort_audit`, with a one-row `summary`, `cells` including eligibility and components, `exclusions`, selected `strata`, `min_cell_n` and a method description. Print and plot methods return the object invisibly.

Examples

```
data <- synthetic_cohorts()
result <- audit_cohorts(data$before, data$after, "district")
print(result)
stopifnot(result$summary$within_change == 0)
stopifnot(result$summary$composition_change == 1200)
```

`synthetic_cohorts` *Invented Snapshots with a Pure Composition Change*

Description

Creates a small deterministic example in which each district has unchanged values but the observed district shares change. No real properties or market observations are included.

Usage

```
synthetic_cohorts()
```

Value

A list with `before` and `after` data frames, each with ten invented observations. Required column names are `listing_id`, `district`, `property_type`, `value`, and `unit`.

Examples

```
example <- synthetic_cohorts()
nrow(example$before)
```

`write_cohort_report` *Write a Portable Local Audit Report*

Description

Writes a self-contained HTML report, summary.csv, cells.csv and exclusions.csv. No network request or automatic upload is performed. Existing files with these names are replaced.

Usage

```
write_cohort_report(x, directory)
```

Arguments

`x`	A `cohort_audit` result.
`directory`	Local destination directory. Created recursively if needed.

Details

HTML text is escaped. CSV character values whose trimmed prefix is a spreadsheet formula marker are prefixed with an apostrophe. These CSVs are presentation copies, not lossless raw data. Reports contain supplied category labels; control access before sharing. Bar lengths show absolute component magnitudes and labels show their signed values.

Value

Invisibly returns the normalized path of the HTML report.

Examples

```
data <- synthetic_cohorts()
result <- audit_cohorts(data$before, data$after, "district")
report <- write_cohort_report(result, tempfile("cohort-report-"))
file.exists(report)
```

Index

audit_cohorts, 2

plot.cohort_audit *(*audit_cohorts*)*, 2
print.cohort_audit *(*audit_cohorts*)*, 2

synthetic_cohorts, 3

write_cohort_report, 4