

Package ‘BKVerify’

September 2, 2026

Type Package

Title Consistency Auditing of Reported Plant Breeding Statistics

Version 0.1.1

Description Audits published or draft plant breeding tables for internal arithmetic consistency. Analysis of variance tables, precision statistics and genetic variability parameters are jointly over-determined by exact algebraic identities; 'BKVerify' recomputes every derivable quantity using rounding-interval arithmetic and reports a value as inconsistent only when no combination of values inside the reported rounding intervals can satisfy the identity. The package deliberately restricts itself to relationships that hold irrespective of which variance-component definition an author adopted, so that flagged results reflect arithmetic inconsistency rather than methodological disagreement. Implemented checks cover analysis of variance internal structure, coefficient of variation, standard error of mean and critical difference, the genetic advance identity of Johnson, Robinson and Comstock (1955) <doi:10.2134/agronj1955.00021962004700070009x>, the relation between genotypic and phenotypic coefficients of variation and broad-sense heritability, and admissibility of reported correlation matrices.

License GPL-3

Encoding UTF-8

Depends R (>= 4.0.0)

Imports stats

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

RoxygenNote 7.3.2

NeedsCompilation no

Author Praveen Kumar [aut, cre]

Maintainer Praveen Kumar <bkpraveenars@gmail.com>

Repository CRAN

Date/Publication 2026-09-02 11:40:08 UTC

Contents

BKVerify-package	2
bk_audit	3
bk_check_anova	4
bk_check_corr	5
bk_check_precision	6
bk_check_variability	7
bk_example	8
bk_interval	9
bk_k	9
Index	11

BKVerify-package	<i>BKVerify: Consistency Auditing of Reported Plant Breeding Statistics</i>
------------------	---

Description

Audits published or draft plant breeding tables for internal arithmetic consistency. Analysis of variance tables, precision statistics and genetic variability parameters are jointly over-determined by exact algebraic identities; **BKVerify** recomputes every derivable quantity using rounding-interval arithmetic and reports a value as inconsistent only when no combination of values inside the reported rounding intervals can satisfy the identity.

Details

The package deliberately restricts itself to relationships that hold irrespective of which variance-component definition an author adopted. Broad-sense heritability is not recomputed from mean squares, because the phenotypic variance is defined in several defensible ways in the plant breeding literature and authors seldom state which they used. The identities used here survive that ambiguity because the phenotypic variance cancels.

The package makes no claim about whether an analysis was appropriate, whether the design was sound, or whether any discrepancy was deliberate. A verdict of "consistent" means only that the printed numbers can be reconciled.

Main functions: [bk_check_variability](#), [bk_check_anova](#), [bk_check_precision](#), [bk_check_corr](#), [bk_audit](#), [bk_example](#), [bk_interval](#), [bk_k](#).

Author(s)

Praveen Kumar <bkpraveenars@gmail.com>

References

Burton, G. W. and DeVane, E. H. (1953). Estimating heritability in tall fescue from replicated clonal material. *Agronomy Journal* 45, 478–481.

Johnson, H. W., Robinson, H. F. and Comstock, R. E. (1955). Estimates of genetic and environmental variability in soybeans. *Agronomy Journal* 47, 314–318.

Nuijten, M. B. and Wicherts, J. M. (2024). Implementing statcheck during peer review is related to a steep decline in statistical-reporting inconsistencies. *Advances in Methods and Practices in Psychological Science* 7.

bk_audit	Combine audit results
----------	-----------------------

Description

Binds the output of several `bk_check_*` calls into a single audit object. The `print` method shows only inconsistent and indeterminate rows, together with a reminder that a "consistent" verdict means the reported values are arithmetically reconcilable, not that the underlying analysis was appropriate.

Usage

```
bk_audit(...)  
  
## S3 method for class 'bk_result'  
print(x, ...)  
  
## S3 method for class 'bk_result'  
summary(object, ...)  
  
## S3 method for class 'bk_result'  
as.data.frame(x, row.names = NULL, optional = FALSE, ...)
```

Arguments

- `...` objects of class `bk_result`, or a single list of them. For the methods, further arguments (ignored).
- `x, object` an object of class `bk_result`.
- `row.names, optional` as in `as.data.frame`.

Value

`bk_audit` returns an object of class `bk_result` carrying an additional module column. The `print` method returns its input invisibly; `summary` returns a verdict-by-test table invisibly; `as.data.frame` returns a plain data frame.

Examples

```
bk_audit(
  bk_check_variability(bk_example("variability")),
  bk_check_anova(bk_example("anova")),
  bk_check_corr(bk_example("corr"))
)
```

bk_check_anova

Audit an analysis of variance table

Description

Checks the internal algebraic structure of a published analysis of variance table: additivity of degrees of freedom and sums of squares, the definition of mean squares ($MS = SS/df$), and the definition of the variance ratio ($F = MS_{source}/MS_{error}$).

Usage

```
bk_check_anova(data, source = "source", df = "df", ss = "ss", ms = "ms",
  f = "f", error_label = "error|residual", total_label = "total",
  decimals = NULL)
```

Arguments

data	a data frame, one row per source of variation.
source, df, ss, ms, f	column names in data. Any of ss, ms, f may be NULL or absent; dependent tests are then skipped.
error_label, total_label	case-insensitive regular expressions identifying the error row and the total row.
decimals	optional integer overriding inferred reporting precision for ss, ms and f.

Details

All comparisons use rounding-interval arithmetic, so a row is flagged only when no combination of values inside the reported rounding intervals can satisfy the identity. Degrees of freedom are treated as exact integers, never as rounded quantities.

Value

An object of class `bk_result`.

See Also

[bk_audit](#), [bk_example](#)

Examples

```
a <- bk_example("anova")
bk_check_anova(a)
```

bk_check_corr

*Audit a reported correlation matrix***Description**

Checks symmetry, unit diagonal, coefficient range and positive semi-definiteness of a reported correlation matrix. A correlation matrix that is not positive semi-definite cannot arise from any data set; published matrices nevertheless fail this because coefficients are computed pairwise on different subsets, transcribed by hand, or assembled from several sources.

Usage

```
bk_check_corr(R, decimals = NULL, symmetry_tol = 1e-08)
```

Arguments

R	a square numeric or character matrix of correlation coefficients. Character input preserves reporting precision and is preferred.
decimals	optional integer overriding the inferred reporting precision.
symmetry_tol	tolerance for the symmetry check. Default 1e-8.

Details

Rounding is handled rigorously. For a symmetric perturbation E with zero diagonal and entries bounded by the rounding half-width δ , Weyl's inequality bounds the eigenvalue movement by $\delta\sqrt{n^2 - n}$. Non-definiteness is reported as "INCONSISTENT" only when $\lambda_{\min}(R) + \delta\sqrt{n^2 - n} < 0$, i.e. when no matrix inside the reported rounding box is positive semi-definite. When the minimum eigenvalue is negative but within the rounding allowance, the verdict is "indeterminate".

Value

An object of class `bk_result`.

See Also

[bk_audit](#), [bk_example](#)

Examples

```
bk_check_corr(bk_example("corr"))
bk_check_corr(bk_example("corr_impossible"))
```

bk_check_precision	<i>Audit precision statistics (CV, SEm, CD)</i>
--------------------	---

Description

Recomputes the coefficient of variation ($CV\% = 100\sqrt{MSe/\bar{x}}$), standard error of the mean ($SEm = \sqrt{MSe/r}$) and critical difference ($CD = t\sqrt{2MSe/r}$) from the error mean square, and additionally checks the relation $CD = t \times SEm \times \sqrt{2}$, which requires only the two printed values and the error degrees of freedom.

Usage

```
bk_check_precision(mse = NULL, r = NULL, grand_mean = NULL,
  df_error = NULL, cv = NULL, sem = NULL, cd = NULL, alpha = 0.05,
  label = NULL, decimals = NULL)
```

Arguments

mse	error mean square. May be NULL if only cd, sem and df_error are available.
r	number of replications.
grand_mean	grand mean of the trait.
df_error	error degrees of freedom (treated as exact).
cv, sem, cd	the reported values to be tested. Any may be NULL.
alpha	significance level used for the critical difference. Default 0.05.
label	optional character vector naming each trait.
decimals	optional integer overriding inferred reporting precision.

Details

All arguments may be vectors of equal length, one element per trait. The $CD = t SEm\sqrt{2}$ relation is useful in practice because many papers print CD and SEm but not the error mean square.

Value

An object of class `bk_result`.

See Also

[bk_audit](#)

Examples

```
bk_check_precision(mse = "10.03", r = 3, grand_mean = "45.16",
  df_error = 38, cv = "7.01", sem = "1.83", cd = "5.24")
```

bk_check_variability *Audit a genetic variability table*

Description

Tests the internal arithmetic of a genotypic/phenotypic coefficient of variation, heritability and genetic advance table using rounding-interval arithmetic. The identities used are $h^2 = (GCV/PCV)^2$, $GAM = Kh^2PCV$, $GAM = KGCV^2/PCV$ and $GCV \leq PCV$.

Usage

```
bk_check_variability(data, trait = "trait", gcv = "gcv", pcv = "pcv",
  h2 = "h2", gam = "gam", k = "auto",
  h2_scale = c("percent", "fraction"), decimals = NULL)
```

Arguments

data	a data frame holding one row per trait.
trait, gcv, pcv, h2, gam	column names in data. Set any of gcv, pcv, h2, gam to NULL if the quantity was not reported; tests requiring it are skipped.
k	either a numeric selection differential, or "auto" (default) to test the standard family returned by <code>bk_k</code> and report which member reconciles each trait.
h2_scale	"percent" (default) if heritability is reported on a 0–100 scale, "fraction" if on a 0–1 scale.
decimals	optional integer or integer vector overriding the inferred reporting precision.

Details

The identities hold irrespective of which phenotypic variance definition the author used, because σ_p^2 cancels. A flagged row therefore indicates arithmetic inconsistency, not methodological disagreement. Values are declared inconsistent only when no combination of values inside the reported rounding intervals satisfies the identity.

Supplying the columns as character vectors is strongly preferred: it preserves trailing zeros and hence the true reporting precision. Numeric input loses them and yields wider intervals, which is conservative.

Because the identities overlap, the pattern of failures localises the fault: a wrong heritability fails `h2_from_cv` and `gam_identity` while `gam_from_cv` passes; a wrong genetic advance fails both genetic advance tests while `h2_from_cv` passes.

Value

An object of class `bk_result`: a data frame with one row per trait per test, carrying the reported interval, the implied interval and a verdict of "consistent", "INCONSISTENT" or "not_applicable".

See Also

[bk_audit](#), [bk_k](#), [bk_example](#)

Examples

```
v <- bk_example("variability")
bk_check_variability(v)
```

bk_example	<i>Example tables for auditing</i>
------------	------------------------------------

Description

Returns small worked examples in the shape the bk_check_* functions expect. Values are stored as character strings so that trailing zeros, and therefore the true reporting precision, are preserved. The tables derive from simulated randomised complete block trials (24 genotypes, 3 replications) rounded to typical published precision.

Usage

```
bk_example(what = c("variability", "anova", "precision", "corr",
                    "corr_impossible"))
```

Arguments

what one of "variability", "anova", "precision", "corr" or "corr_impossible".

Details

The "variability" table carries two planted errors, useful for teaching error localisation:

Grains per spike heritability inflated by 4 points. The h2_from_cv and gam_identity tests fail while gam_from_cv passes, localising the fault to the heritability column.

Grain yield per plant genetic advance inflated by 15 percent. Here h2_from_cv passes and both genetic advance tests fail, localising the fault to the genetic advance column.

Value

A data frame, matrix or list, depending on what.

Examples

```
bk_example("variability")
bk_check_variability(bk_example("variability"))
```

bk_interval	<i>Rounding interval of a reported value</i>
-------------	--

Description

Converts reported values into the closed intervals they actually assert. A value printed to d decimals implies a half-width of $0.5 \times 10^{(-d)}$. Every check in **BKVerify** is decided on these intervals, never on point values.

Usage

```
bk_interval(x, decimals = NULL)
```

Arguments

x	numeric or character vector of reported values. Character input is preferred: it preserves trailing zeros and therefore the true reporting precision.
decimals	optional integer. If supplied, overrides the decimal count inferred from x. Recycled to the length of x.

Value

A two-column matrix with columns lo and hi.

Examples

```
bk_interval(c("11.04", "11.76"))
bk_interval(88.14, decimals = 2)
```

bk_k	<i>Selection differential constants (K) used in genetic advance</i>
------	---

Description

Genetic advance is computed as $GA = K\sigma_p h^2$, where K is the standardised selection differential for the chosen selection intensity (Johnson, Robinson and Comstock 1955). Authors overwhelmingly report the conventional rounded values (2.64, 2.06, 1.76, 1.40) rather than exact normal-curve quantiles, so those conventional values are what **BKVerify** tests against.

Usage

```
bk_k(intensity = NULL)
```

Arguments

intensity	optional character; one of "1%", "5%", "10%", "20%". If NULL, the full named vector is returned.
-----------	--

Details

Because a paper that omits its selection intensity is indistinguishable from one that used a different intensity, `k = "auto"` in [bk_check_variability](#) tests the whole family and reports which member reconciles the table.

Value

A named numeric vector, or a single value if `intensity` is given.

References

Johnson, H. W., Robinson, H. F. and Comstock, R. E. (1955). Estimates of genetic and environmental variability in soybeans. *Agronomy Journal* 47, 314–318.

Examples

```
bk_k()  
bk_k("5%")
```

Index

`as.data.frame`, [3](#)
`as.data.frame.bk_result (bk_audit)`, [3](#)

`bk_audit`, [2](#), [3](#), [4–6](#), [8](#)
`bk_check_anova`, [2](#), [4](#)
`bk_check_corr`, [2](#), [5](#)
`bk_check_precision`, [2](#), [6](#)
`bk_check_variability`, [2](#), [7](#), [10](#)
`bk_example`, [2](#), [4](#), [5](#), [8](#), [8](#)
`bk_interval`, [2](#), [9](#)
`bk_k`, [2](#), [7](#), [8](#), [9](#)
`BKVerify (BKVerify-package)`, [2](#)
`BKVerify-package`, [2](#)

`print.bk_result (bk_audit)`, [3](#)

`summary.bk_result (bk_audit)`, [3](#)