

# Package ‘ReplicationSuccess’

April 3, 2023

**Type** Package

**Title** Design and Analysis of Replication Studies

**Version** 1.3.1

**Date** 2023-04-03

**Description** Provides utilities for the design and analysis of replication studies. Features both traditional methods based on statistical significance and more recent methods such as the sceptical p-value; Held L. (2020) <[doi:10.1111/rssa.12493](https://doi.org/10.1111/rssa.12493)>, Held et al. (2022) <[doi:10.1214/21-AOAS1502](https://doi.org/10.1214/21-AOAS1502)>, Micheloud et al. (2023) <[arXiv:2207.00464](https://arxiv.org/abs/2207.00464)>. Also provides related methods including the harmonic mean chi-squared test; Held, L. (2020) <[doi:10.1111/rssc.12410](https://doi.org/10.1111/rssc.12410)>, and intrinsic credibility; Held, L. (2019) <[doi:10.1098/rsos.181534](https://doi.org/10.1098/rsos.181534)>. Contains datasets from five large-scale replication projects.

**License** GPL (>= 2)

**URL** <https://crsuzh.github.io/ReplicationSuccess/>

**BugReports** <https://github.com/crsuzh/ReplicationSuccess/issues/>

**Imports** stats

**Suggests** knitr, roxygen2, testthat

**VignetteBuilder** knitr

**Encoding** UTF-8

**LazyData** true

**NeedsCompilation** no

**RoxygenNote** 7.2.3

**Author** Leonhard Held [aut] (<<https://orcid.org/0000-0002-8686-5325>>),  
Samuel Pawel [aut] (<<https://orcid.org/0000-0003-2779-320X>>),  
Charlotte Micheloud [aut] (<<https://orcid.org/0000-0002-4995-4505>>),  
Florian Gerber [aut] (<<https://orcid.org/0000-0001-8545-5263>>),  
Felix Hofmann [cre] (<<https://orcid.org/0000-0002-3891-6239>>)

**Maintainer** Felix Hofmann <[felix.hofmann2@uzh.ch](mailto:felix.hofmann2@uzh.ch)>

**Repository** CRAN

**Date/Publication** 2023-04-03 13:20:05 UTC

**R topics documented:**

ci2se . . . . .	2
effectSizeReplicationSuccess . . . . .	4
effectSizeSignificance . . . . .	6
hMeanChiSq . . . . .	7
levelSceptical . . . . .	10
pBox . . . . .	11
pIntrinsic . . . . .	13
powerReplicationSuccess . . . . .	14
powerSignificance . . . . .	16
powerSignificanceInterim . . . . .	18
PPpSceptical . . . . .	20
predictionInterval . . . . .	22
pReplicate . . . . .	24
protzko2020 . . . . .	25
pSceptical . . . . .	27
pvalueBound . . . . .	29
Qtest . . . . .	30
RProjects . . . . .	31
sampleSizeReplicationSuccess . . . . .	33
sampleSizeSignificance . . . . .	35
SSRP . . . . .	37
T1EpSceptical . . . . .	39
thresholdIntrinsic . . . . .	40
<b>Index</b>	<b>42</b>

---

ci2se	<i>Convert between estimates, z-values, p-values, and confidence intervals</i>
-------	--

---

**Description**

Convert between estimates, z-values, p-values, and confidence intervals

**Usage**

```
ci2se(lower, upper, conf.level = 0.95, ratio = FALSE)
```

```
ci2estimate(lower, upper, ratio = FALSE, antilog = FALSE)
```

```
ci2z(lower, upper, conf.level = 0.95, ratio = FALSE)
```

```
ci2p(
  lower,
  upper,
  conf.level = 0.95,
```

```

    ratio = FALSE,
    alternative = c("two.sided", "one.sided", "less", "greater")
  )

z2p(z, alternative = c("two.sided", "one.sided", "less", "greater"))

p2z(p, alternative = c("two.sided", "one.sided", "less", "greater"))

```

### Arguments

lower	Numeric vector of lower confidence interval bounds.
upper	Numeric vector of upper confidence interval bounds.
conf.level	The confidence level of the confidence intervals. Default is 0.95.
ratio	Indicates whether the confidence interval is for a ratio, e.g. an odds ratio, relative risk or hazard ratio. If TRUE, the standard error of the log ratio is computed. Defaults to FALSE.
antilog	Indicates whether the estimate is reported on the ratio scale. Only applies if ratio = TRUE. Defaults to FALSE.
alternative	Direction of the alternative of the p-value. Either "two.sided" (default), "one.sided", "less", or "greater". If "one.sided" or "two.sided" is specified, the z-value is assumed to be positive.
z	Numeric vector of z-values.
p	Numeric vector of p-values.

### Details

z2p is the vectorized version of the internal function .z2p\_. [Vectorize](#) is used to vectorize the function.

p2z is the vectorized version of the internal function .p2z\_. [Vectorize](#) is used to vectorize the function.

### Value

ci2se returns a numeric vector of standard errors.

ci2estimate returns a numeric vector of parameter estimates.

ci2z returns a numeric vector of z-values.

ci2p returns a numeric vector of p-values.

z2p returns a numeric vector of p-values.

p2z returns a numeric vector of z-values.

### Examples

```

ci2se(lower = 1, upper = 3)
ci2se(lower = 1, upper = 3, ratio = TRUE)
ci2se(lower = 1, upper = 3, conf.level = 0.9)

```

```

ci2estimate(lower = 1, upper = 3)
ci2estimate(lower = 1, upper = 3, ratio = TRUE)
ci2estimate(lower = 1, upper = 3, ratio = TRUE, antilog = TRUE)

ci2z(lower = 1, upper = 3)
ci2z(lower = 1, upper = 3, ratio = TRUE)
ci2z(lower = 1, upper = 3, conf.level = 0.9)

ci2p(lower = 1, upper = 3)
ci2p(lower = 1, upper = 3, alternative = "one.sided")

z2p(z = c(1, 2, 5))
z2p(z = c(1, 2, 5), alternative = "less")
z2p(z = c(1, 2, 5), alternative = "greater")
z <- seq(-3, 3, by = 0.01)
plot(z, z2p(z), type = "l", xlab = "z", ylab = "p", ylim = c(0, 1))
lines(z, z2p(z, alternative = "greater"), lty = 2)
legend("topright", c("two-sided", "greater"), lty = c(1, 2), bty = "n")

p2z(p = c(0.005, 0.01, 0.05))
p2z(p = c(0.005, 0.01, 0.05), alternative = "greater")
p2z(p = c(0.005, 0.01, 0.05), alternative = "less")
p <- seq(0.001, 0.05, 0.0001)
plot(p, p2z(p), type = "l", ylim = c(0, 3.5), ylab = "z")
lines(p, p2z(p, alternative = "greater"), lty = 2)
legend("bottomleft", c("two-sided", "greater"), lty = c(1, 2), bty = "n")

```

---

effectSizeReplicationSuccess

*Computes the minimum relative effect size to achieve replication success with the sceptical p-value*

---

## Description

The minimum relative effect size (replication to original) to achieve replication success with the sceptical p-value is computed based on the result of the original study and the corresponding variance ratio.

## Usage

```

effectSizeReplicationSuccess(
  zo,
  c = 1,
  level = 0.025,
  alternative = c("one.sided", "two.sided"),
  type = c("golden", "nominal", "controlled")
)

```

**Arguments**

zo	Numeric vector of z-values from original studies.
c	Numeric vector of variance ratios of the original and replication effect estimates. This is usually the ratio of the sample size of the replication study to the sample size of the original study.
level	Threshold for the calibrated sceptical p-value. Default is 0.025.
alternative	Specifies if level is "one.sided" (default) or "two.sided". If "one.sided", then effect size calculations are based on a one-sided assessment of replication success in the direction of the original effect estimate.
type	Type of recalibration. Can be either "golden" (default), "nominal" (no recalibration), or "controlled". "golden" ensures that for an original study just significant at the specified level, replication success is only possible for replication effect estimates larger than the original one. "controlled" ensures exact overall Type-I error control at level $\text{level}^2$ .

**Details**

effectSizeReplicationSuccess is the vectorized version of the internal function `.effectSizeReplicationSuccess_`. [Vectorize](#) is used to vectorize the function.

**Value**

The minimum relative effect size to achieve replication success with the sceptical p-value.

**Author(s)**

Leonhard Held, Charlotte Micheloud, Samuel Pawel, Florian Gerber

**References**

- Held, L., Micheloud, C., Pawel, S. (2022). The assessment of replication success based on relative effect size. *The Annals of Applied Statistics*. 16:706-720. [doi:10.1214/21AOAS1502](https://doi.org/10.1214/21AOAS1502)
- Micheloud, C., Balabdaoui, F., Held, L. (2023). Beyond the two-trials rule: Type-I error control and sample size planning with the sceptical p-value. <https://arxiv.org/abs/2207.00464>

**See Also**

[sampleSizeReplicationSuccess](#), [levelSceptical](#)

**Examples**

```
po <- c(0.001, 0.002, 0.01, 0.02, 0.025)
zo <- p2z(po, alternative = "one.sided")

effectSizeReplicationSuccess(zo = zo, c = 1, level = 0.025,
                             alternative = "one.sided", type = "golden")

effectSizeReplicationSuccess(zo = zo, c = 10, level = 0.025,
```

```

                                alternative = "one.sided", type = "golden")
effectSizeReplicationSuccess(zo = zo, c = 10, level = 0.025,
                                alternative = "one.sided", type = "controlled")
effectSizeReplicationSuccess(zo = zo, c = 2, level = 0.025,
                                alternative = "one.sided", type = "nominal")

effectSizeReplicationSuccess(zo = zo, c = 2, level = 0.05,
                                alternative = "two.sided", type = "nominal")

```

---

effectSizeSignificance

*Computes the minimum relative effect size to achieve significance of the replication study*

---

### Description

The minimum relative effect size (replication to original) to achieve significance of the replication study is computed based on the result of the original study and the corresponding variance ratio.

### Usage

```

effectSizeSignificance(
  zo,
  c = 1,
  level = 0.025,
  alternative = c("one.sided", "two.sided")
)

```

### Arguments

zo	Numeric vector of z-values from original studies.
c	Numeric vector of variance ratios of the original and replication effect estimates. This is usually the ratio of the sample size of the replication study to the sample size of the original study.
level	Significance level. Default is 0.025.
alternative	Specifies if the significance level is "one.sided" (default) or "two.sided". If the significance level is one-sided, then effect size calculations are based on a one-sided assessment of significance in the direction of the original effect estimate.

### Details

effectSizeSignificance is the vectorized version of the internal function .effectSizeSignificance\_. [Vectorize](#) is used to vectorize the function.

### Value

The minimum relative effect size to achieve significance in the replication study.

**Author(s)**

Charlotte Micheloud, Samuel Pawel, Florian Gerber

**References**

Held, L., Micheloud, C., Pawel, S. (2022). The assessment of replication success based on relative effect size. *The Annals of Applied Statistics*. 16:706-720. doi:10.1214/21AOAS1502

**See Also**

[effectSizeReplicationSuccess](#)

**Examples**

```
po <- c(0.001, 0.002, 0.01, 0.02, 0.025)
zo <- p2z(po, alternative = "one.sided")

effectSizeSignificance(zo = zo, c = 1, level = 0.025,
                      alternative = "one.sided")

effectSizeSignificance(zo = zo, c = 1, level = 0.05,
                      alternative = "two.sided")

effectSizeSignificance(zo = zo, c = 50, level = 0.025,
                      alternative = "one.sided")
```

---

hMeanChiSq

*harmonic mean chi-squared test*

---

**Description**

p-values and confidence intervals from the harmonic mean chi-squared test.

**Usage**

```
hMeanChiSq(
  z,
  w = rep(1, length(z)),
  alternative = c("greater", "less", "two.sided", "none"),
  bound = FALSE
)

hMeanChiSqMu(
  thetahat,
  se,
  w = rep(1, length(thetahat)),
  mu = 0,
  alternative = c("greater", "less", "two.sided", "none"),
```

```

    bound = FALSE
  )

hMeanChiSqCI(
  thetahat,
  se,
  w = rep(1, length(thetahat)),
  alternative = c("two.sided", "greater", "less", "none"),
  conf.level = 0.95
)

```

### Arguments

<code>z</code>	Numeric vector of z-values.
<code>w</code>	Numeric vector of weights.
<code>alternative</code>	Either "greater" (default), "less", "two.sided", or "none". Specifies the alternative to be considered in the computation of the p-value.
<code>bound</code>	If FALSE (default), p-values that cannot be computed are reported as NaN. If TRUE, they are reported as "> bound".
<code>thetahat</code>	Numeric vector of parameter estimates.
<code>se</code>	Numeric vector of standard errors.
<code>mu</code>	The null hypothesis value. Defaults to 0.
<code>conf.level</code>	Numeric vector specifying the conf.level of the confidence interval. Defaults to 0.95. summarize the gamma values, i.e., the local minima of the p-value function between the thetahats. Defaults is a vector of 1s.

### Value

`hMeanChiSq`: returns the p-values from the harmonic mean chi-squared test based on the study-specific z-values.

`hMeanChiSqMu`: returns the p-value from the harmonic mean chi-squared test based on study-specific estimates and standard errors.

`hMeanChiSqCI`: returns a list containing confidence interval(s) obtained by inverting the harmonic mean chi-squared test based on study-specific estimates and standard errors. The list contains:

<code>CI</code>	Confidence interval(s).
-----------------	-------------------------

If the `alternative` is "none", the list also contains:

<code>gamma</code>	Local minima of the p-value function between the thetahats.
--------------------	---

### Author(s)

Leonhard Held, Florian Gerber



## References

Held, L. (2020). The harmonic mean chi-squared test to substantiate scientific findings. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **69**, 697-708. doi:10.1111/rssc.12410

## Examples

```
## Example from Fisher (1999) as discussed in Held (2020)
pvalues <- c(0.0245, 0.1305, 0.00025, 0.2575, 0.128)
lower <- c(0.04, 0.21, 0.12, 0.07, 0.41)
upper <- c(1.14, 1.54, 0.60, 3.75, 1.27)
se <- ci2se(lower = lower, upper = upper, ratio = TRUE)
thetahat <- ci2estimate(lower = lower, upper = upper, ratio = TRUE)

## hMeanChiSq() -----
hMeanChiSq(z = p2z(p = pvalues, alternative = "less"),
           alternative = "less")
hMeanChiSq(z = p2z(p = pvalues, alternative = "less"),
           alternative = "two.sided")
hMeanChiSq(z = p2z(p = pvalues, alternative = "less"),
           alternative = "none")

hMeanChiSq(z = p2z(p = pvalues, alternative = "less"),
           w = 1 / se^2, alternative = "less")
hMeanChiSq(z = p2z(p = pvalues, alternative = "less"),
           w = 1 / se^2, alternative = "two.sided")
hMeanChiSq(z = p2z(p = pvalues, alternative = "less"),
           w = 1 / se^2, alternative = "none")

## hMeanChiSqMu() -----
hMeanChiSqMu(thetahat = thetahat, se = se, alternative = "two.sided")
hMeanChiSqMu(thetahat = thetahat, se = se, w = 1 / se^2,
             alternative = "two.sided")
hMeanChiSqMu(thetahat = thetahat, se = se, alternative = "two.sided",
             mu = -0.1)

## hMeanChiSqCI() -----
## two-sided
CI1 <- hMeanChiSqCI(thetahat = thetahat, se = se, w = 1 / se^2,
                  alternative = "two.sided")
CI2 <- hMeanChiSqCI(thetahat = thetahat, se = se, w = 1 / se^2,
                  alternative = "two.sided", conf.level = 0.99875)

## one-sided
CI1b <- hMeanChiSqCI(thetahat = thetahat, se = se, w = 1 / se^2,
                   alternative = "less", conf.level = 0.975)
CI2b <- hMeanChiSqCI(thetahat = thetahat, se = se, w = 1 / se^2,
                   alternative = "less", conf.level = 1 - 0.025^2)

## confidence intervals on hazard ratio scale
print(exp(CI1$CI), digits = 2)
print(exp(CI2$CI), digits = 2)
print(exp(CI1b$CI), digits = 2)
```

```

print(exp(CI2b$CI), digits = 2)

## example with confidence region consisting of disjunct intervals
thetahat2 <- c(-3.7, 2.1, 2.5)
se2 <- c(1.5, 2.2, 3.1)
conf.level <- 0.95; alpha <- 1 - conf.level
muSeq <- seq(-7, 6, length.out = 1000)
pValueSeq <- hMeanChiSqMu(thetahat = thetahat2, se = se2,
                          alternative = "none", mu = muSeq)
(hm <- hMeanChiSqCI(thetahat = thetahat2, se = se2, alternative = "none"))

plot(x = muSeq, y = pValueSeq, type = "l", panel.first = grid(lty = 1),
     xlab = expression(mu), ylab = "p-value")
abline(v = thetahat2, h = alpha, lty = 2)
arrows(x0 = hm$CI[, 1], x1 = hm$CI[, 2], y0 = alpha,
       y1 = alpha, col = "darkgreen", lwd = 3, angle = 90, code = 3)
points(hm$gamma, col = "red", pch = 19, cex = 2)

```

---

levelSceptical

*Computes the replication success level*


---

## Description

The replication success level is computed based on the specified alternative and recalibration type.

## Usage

```

levelSceptical(
  level,
  c = NA,
  alternative = c("one.sided", "two.sided"),
  type = c("golden", "nominal", "controlled")
)

```

## Arguments

level	Threshold for the calibrated sceptical p-value. Default is 0.025.
c	The variance ratio. Only required when type = "controlled".
alternative	Specifies if level is "one.sided" (default) or "two.sided". If "one-sided", then a one-sided replication success level is computed.
type	Type of recalibration. Can be either "golden" (default), "nominal" (no recalibration), or "controlled". "golden" ensures that for an original study just significant at the specified level, replication success is only possible for replication effect estimates larger than the original one. "controlled" ensures exact overall Type-I error control at level $\text{level}^2$ .

## Details

`levelSceptical` is the vectorized version of the internal function `.levelSceptical_`. [Vectorize](#) is used to vectorize the function.

## Value

Replication success levels

## Author(s)

Leonhard Held

## References

- Held, L. (2020). A new standard for the analysis and design of replication studies (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **183**, 431-448. doi:10.1111/rssa.12493
- Held, L. (2020). The harmonic mean chi-squared test to substantiate scientific findings. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **69**, 697-708. doi:10.1111/rssc.12410
- Held, L., Micheloud, C., Pawel, S. (2022). The assessment of replication success based on relative effect size. *The Annals of Applied Statistics*, **16**, 706-720. doi:10.1214/21AOAS1502
- Micheloud, C., Balabdaoui, F., Held, L. (2023). Beyond the two-trials rule: Type-I error control and sample size planning with the sceptical p-value. <https://arxiv.org/abs/2207.00464>

## Examples

```
levelSceptical(level = 0.025, alternative = "one.sided", type = "nominal")
levelSceptical(level = 0.025, alternative = "one.sided", type = "controlled", c = 1)
levelSceptical(level = 0.025, alternative = "one.sided", type = "golden")
```

---

pBox

*Computes Box's tail probability*

---

## Description

`pBox` computes Box's tail probabilities based on the z-values of the original and the replication study, the corresponding variance ratio, and the significance level.

## Usage

```
pBox(zo, zr, c, level = 0.05, alternative = c("two.sided", "one.sided"))
zBox(zo, zr, c, level = 0.05, alternative = c("two.sided", "one.sided"))
```

**Arguments**

zo	Numeric vector of z-values from the original studies.
zr	Numeric vector of z-values from replication studies.
c	Numeric vector of variance ratios of the original and replication effect estimates. This is usually the ratio of the sample size of the replication study to the sample size of the original study.
level	Numeric vector of significance levels. Default is 0.05.
alternative	Either "two.sided" (default) or "one.sided". Specifies whether two-sided or one-sided Box's tail probabilities are computed.

**Details**

pBox quantifies the conflict between the sceptical prior that would render the original study non-significant and the result from the replication study. If the original study was not significant at level level, the sceptical prior does not exist and pBox cannot be calculated.

**Value**

pBox returns Box's tail probabilities.

zBox returns the z-values used in pBox.

**Author(s)**

Leonhard Held

**References**

Box, G.E.P. (1980). Sampling and Bayes' inference in scientific modelling and robustness (with discussion). *Journal of the Royal Statistical Society, Series A*, **143**, 383-430.

Held, L. (2020). A new standard for the analysis and design of replication studies (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **183**, 431-448. doi:[10.1111/rssa.12493](https://doi.org/10.1111/rssa.12493)

**Examples**

```
pBox(zo = p2z(0.01), zr = p2z(0.02), c = 2)
pBox(zo = p2z(0.02), zr = p2z(0.01), c = 1/2)
pBox(zo = p2z(0.02, alternative = "one.sided"),
      zr = p2z(0.01, alternative = "one.sided"),
      c = 1/2, alternative = "one.sided")
```

---

pIntrinsic                      *Computes the p-value for intrinsic credibility*

---

### Description

Computes the p-value for intrinsic credibility

### Usage

```
pIntrinsic(  
  p = z2p(z, alternative = alternative),  
  z = NULL,  
  alternative = c("two.sided", "one.sided", "less", "greater"),  
  type = c("Held", "Matthews")  
)
```

### Arguments

p	numeric vector of p-values.
z	numeric vector of z-values. Default is NULL.
alternative	Either "two.sided" (default) or "one.sided". Specifies if the p-value is two-sided or one-sided. If the p-value is one-sided, then a one-sided p-value for intrinsic credibility is computed.
type	Type of intrinsic p-value. Default is "Held" as in Held (2019). The other option is "Matthews" as in Matthews (2018).

### Value

p-values for intrinsic credibility.

### Author(s)

Leonhard Held

### References

Matthews, R. A. J. (2018). Beyond 'significance': principles and practice of the analysis of credibility. *Royal Society Open Science*, **5**, 171047. doi:10.1098/rsos.171047

Held, L. (2019). The assessment of intrinsic credibility and a new argument for  $p < 0.005$ . *Royal Society Open Science*, **6**, 181534. doi:10.1098/rsos.181534

**Examples**

```
p <- c(0.005, 0.01, 0.05)
pIntrinsic(p = p)
pIntrinsic(p = p, type = "Matthews")
pIntrinsic(p = p, alternative = "one.sided")
pIntrinsic(p = p, alternative = "one.sided", type = "Matthews")

pIntrinsic(z = 2)
```

---

powerReplicationSuccess

*Computes the power for replication success with the sceptical p-value*

---

**Description**

Computes the power for replication success with the sceptical p-value based on the result of the original study, the corresponding variance ratio, and the design prior.

**Usage**

```
powerReplicationSuccess(
  zo,
  c = 1,
  level = 0.025,
  designPrior = c("conditional", "predictive", "EB"),
  alternative = c("one.sided", "two.sided"),
  type = c("golden", "nominal", "controlled"),
  shrinkage = 0,
  h = 0,
  strict = FALSE
)
```

**Arguments**

zo	Numeric vector of z-values from original studies.
c	Numeric vector of variance ratios of the original and replication effect estimates. This is usually the ratio of the sample size of the replication study to the sample size of the original study.
level	Threshold for the calibrated sceptical p-value. Default is 0.025.
designPrior	Either "conditional" (default), "predictive", or "EB". If "EB", the power is computed under a predictive distribution, where the contribution of the original study is shrunk towards zero based on the evidence in the original study (with an empirical Bayes shrinkage estimator).
alternative	Specifies if level is "one.sided" (default) or "two.sided". If "one.sided" then power calculations are based on a one-sided assessment of replication success in the direction of the original effect estimates.

type	Type of recalibration. Can be either "golden" (default), "nominal" (no recalibration), or "controlled". "golden" ensures that for an original study just significant at the specified level, replication success is only possible for replication effect estimates larger than the original one. "controlled" ensures exact overall Type-I error control at level $\text{level}^2$ .
shrinkage	Numeric vector with values in $[0,1)$ . Defaults to 0. Specifies the shrinkage of the original effect estimate towards zero, e.g., the effect is shrunk by a factor of 25% for <code>shrinkage = 0.25</code> . Is only taken into account if the <code>designPrior</code> is "conditional" or "predictive".
h	Numeric vector of relative heterogeneity variances i.e., the ratios of the heterogeneity variance to the variance of the original effect estimate. Default is 0 (no heterogeneity). Is only taken into account when <code>designPrior = "predictive"</code> or <code>designPrior = "EB"</code> .
strict	Logical vector indicating whether the probability for replication success in the opposite direction of the original effect estimate should also be taken into account. Default is FALSE. Only taken into account when <code>alternative = "two.sided"</code> .

### Details

`powerReplicationSuccess` is the vectorized version of the internal function `.powerReplicationSuccess_`. [Vectorize](#) is used to vectorize the function.

### Value

The power for replication success with the sceptical p-value

### Author(s)

Leonhard Held, Charlotte Micheloud, Samuel Pawel

### References

- Held, L. (2020). A new standard for the analysis and design of replication studies (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **183**, 431-448. doi:[10.1111/rssa.12493](https://doi.org/10.1111/rssa.12493)
- Held, L., Micheloud, C., Pawel, S. (2022). The assessment of replication success based on relative effect size. *The Annals of Applied Statistics*. 16:706-720. doi:[10.1214/21AOAS1502](https://doi.org/10.1214/21AOAS1502)
- Micheloud, C., Balabdaoui, F., Held, L. (2023). Beyond the two-trials rule: Type-I error control and sample size planning with the sceptical p-value. <https://arxiv.org/abs/2207.00464>

### See Also

[sampleSizeReplicationSuccess](#), [pSceptical](#), [levelSceptical](#)

**Examples**

```
## larger sample size in replication (c > 1)
powerReplicationSuccess(zo = p2z(0.005), c = 2, level = 0.025, designPrior = "conditional")
powerReplicationSuccess(zo = p2z(0.005), c = 2, level = 0.025, designPrior = "predictive")

## smaller sample size in replication (c < 1)
powerReplicationSuccess(zo = p2z(0.005), c = 1/2, level = 0.025, designPrior = "conditional")
powerReplicationSuccess(zo = p2z(0.005), c = 1/2, level = 0.025, designPrior = "predictive")

powerReplicationSuccess(zo = p2z(0.00005), c = 2, level = 0.05,
                        alternative = "two.sided", strict = TRUE, shrinkage = 0.9)
powerReplicationSuccess(zo = p2z(0.00005), c = 2, level = 0.05,
                        alternative = "two.sided", strict = FALSE, shrinkage = 0.9)
```

---

powerSignificance      *Computes the power for significance*

---

**Description**

The power for significance is computed based on the result of the original study, the corresponding variance ratio, and the design prior.

**Usage**

```
powerSignificance(
  zo,
  c = 1,
  level = 0.025,
  designPrior = c("conditional", "predictive", "EB"),
  alternative = c("one.sided", "two.sided"),
  h = 0,
  shrinkage = 0,
  strict = FALSE
)
```

**Arguments**

zo	Numeric vector of z-values from original studies.
c	Numeric vector of variance ratios of the original and replication effect estimates. This is usually the ratio of the sample size of the replication study to the sample size of the original study.
level	Significance level. Default is 0.025.
designPrior	Either "conditional" (default), "predictive", or "EB". If "EB", the power is computed under a predictive distribution, where the contribution of the original study is shrunken towards zero based on the evidence in the original study (with an empirical Bayes shrinkage estimator).



alternative	Either "one.sided" (default) or "two.sided". Specifies if the significance level is one-sided or two-sided. If the significance level is one-sided, then power calculations are based on a one-sided assessment of significance in the direction of the original effect estimates.
h	The relative between-study heterogeneity, i.e., the ratio of the heterogeneity variance to the variance of the original effect estimate. Default is 0 (no heterogeneity). Is only taken into account when designPrior = "predictive" or designPrior = "EB".
shrinkage	Numeric vector with values in [0,1). Defaults to 0. Specifies the shrinkage of the original effect estimate towards zero, e.g., the effect is shrunk by a factor of 25% for shrinkage = 0.25. Is only taken into account if the designPrior is "conditional" or "predictive".
strict	Logical vector indicating whether the probability for significance in the opposite direction of the original effect estimate should also be taken into account. Default is FALSE. Only taken into account when alternative = "two.sided".

### Details

powerSignificance is the vectorized version of the internal function `.powerSignificance_`. [Vectorize](#) is used to vectorize the function.

### Value

The probability that a replication study yields a significant effect estimate in the specified direction.

### Author(s)

Leonhard Held, Samuel Pawel, Charlotte Micheloud, Florian Gerber

### References

- Goodman, S. N. (1992). A comment on replication, p-values and evidence, *Statistics in Medicine*, **11**, 875–879. doi:10.1002/sim.4780110705
- Senn, S. (2002). Letter to the Editor, *Statistics in Medicine*, **21**, 2437–2444.
- Held, L. (2020). A new standard for the analysis and design of replication studies (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **183**, 431-448. doi:10.1111/rssa.12493
- Pawel, S., Held, L. (2020). Probabilistic forecasting of replication studies. *PLoS ONE*. **15**, e0231416. doi:10.1371/journal.pone.0231416
- Held, L., Micheloud, C., Pawel, S. (2022). The assessment of replication success based on relative effect size. *The Annals of Applied Statistics*. 16:706-720. doi:10.1214/21AOAS1502
- Micheloud, C., Held, L. (2022). Power Calculations for Replication Studies. *Statistical Science*. 37:369-379. doi:10.1214/21STS828

### See Also

[sampleSizeSignificance](#), [powerSignificanceInterim](#)

**Examples**

```

powerSignificance(zo = p2z(0.005), c = 2)
powerSignificance(zo = p2z(0.005), c = 2, designPrior = "predictive")
powerSignificance(zo = p2z(0.005), c = 2, alternative = "two.sided")
powerSignificance(zo = -3, c = 2, designPrior = "predictive",
                  alternative = "one.sided")
powerSignificance(zo = p2z(0.005), c = 1/2)
powerSignificance(zo = p2z(0.005), c = 1/2, designPrior = "predictive")
powerSignificance(zo = p2z(0.005), c = 1/2, alternative = "two.sided")
powerSignificance(zo = p2z(0.005), c = 1/2, designPrior = "predictive",
                  alternative = "two.sided")
powerSignificance(zo = p2z(0.005), c = 1/2, designPrior = "predictive",
                  alternative = "one.sided", h = 0.5, shrinkage = 0.5)
powerSignificance(zo = p2z(0.005), c = 1/2, designPrior = "EB",
                  alternative = "two.sided", h = 0.5)

# power as function of original p-value
po <- seq(0.0001, 0.06, 0.0001)
plot(po, powerSignificance(zo = p2z(po), designPrior = "conditional"),
     type = "l", ylim = c(0, 1), lwd = 1.5, las = 1, ylab = "Power",
     xlab = expression(italic(p)[o]))
lines(po, powerSignificance(zo = p2z(po), designPrior = "predictive"),
      lwd = 2, lty = 2)
lines(po, powerSignificance(zo = p2z(po), designPrior = "EB"),
      lwd = 1.5, lty = 3)
legend("topright", legend = c("conditional", "predictive", "EB"),
      title = "Design prior", lty = c(1, 2, 3), lwd = 1.5, bty = "n")

```

---

powerSignificanceInterim

*Interim power of a replication study*

---

**Description**

Computes the power of a replication study taking into account data from an interim analysis.

**Usage**

```

powerSignificanceInterim(
  zo,
  zi,
  c = 1,
  f = 1/2,
  level = 0.025,
  designPrior = c("conditional", "informed predictive", "predictive"),
  analysisPrior = c("flat", "original"),
  alternative = c("one.sided", "two.sided"),
  shrinkage = 0
)

```

**Arguments**

zo	Numeric vector of z-values from original studies.
zi	Numeric vector of z-values from interim analyses of replication studies.
c	Numeric vector of variance ratios of the original and replication effect estimates. This is usually the ratio of the sample size of the replication study to the sample size of the original study. Default is 1.
f	Fraction of the replication study already completed. Default is 0.5.
level	Significance level. Default is 0.025.
designPrior	Either "conditional" (default), "informed predictive", or "predictive". "informed predictive" refers to an informative normal prior coming from the original study. "predictive" refers to a flat prior.
analysisPrior	Either "flat" (default) or "original".
alternative	Either "one.sided" (default) or "two.sided". Specifies if the significance level is one-sided or two-sided.
shrinkage	Numeric vector with values in [0,1). Defaults to 0. Specifies the shrinkage of the original effect estimate towards zero, e.g., the effect is shrunken by a factor of 25% for shrinkage=0.25.

**Details**

This is an extension of `powerSignificance()` and adapts the ‘interim power’ from section 6.6.3 of Spiegelhalter et al. (2004) to the setting of replication studies.

`powerSignificanceInterim` is the vectorized version of `.powerSignificanceInterim_`. [Vectorize](#) is used to vectorize the function.

**Value**

The probability of statistical significance in the specified direction at the end of the replication study given the data collected so far in the replication study.

**Author(s)**

Charlotte Micheloud

**References**

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). Bayesian Approaches to Clinical Trials and Health-Care Evaluation, volume 13. John Wiley & Sons

Micheloud, C., Held, L. (2022). Power Calculations for Replication Studies. *Statistical Science*, 37, 369-379. doi:10.1214/21STS828

**See Also**

[sampleSizeSignificance](#), [powerSignificance](#)

**Examples**

```

powerSignificanceInterim(zo = 2, zi = 2, c = 1, f = 1/2,
  designPrior = "conditional",
  analysisPrior = "flat")

powerSignificanceInterim(zo = 2, zi = 2, c = 1, f = 1/2,
  designPrior = "informed predictive",
  analysisPrior = "flat")

powerSignificanceInterim(zo = 2, zi = 2, c = 1, f = 1/2,
  designPrior = "predictive",
  analysisPrior = "flat")

powerSignificanceInterim(zo = 2, zi = -2, c = 1, f = 1/2,
  designPrior = "conditional",
  analysisPrior = "flat")

powerSignificanceInterim(zo = 2, zi = 2, c = 1, f = 1/2,
  designPrior = "conditional",
  analysisPrior = "flat",
  shrinkage = 0.25)

```

---

 PPpSceptical

---

*Compute project power of the sceptical p-value*


---

**Description**

The project power of the sceptical p-value is computed for a specified level, the relative variance, significance level and power for a standard significance test of the original study, and the alternative hypothesis.

**Usage**

```

PPpSceptical(
  level,
  c,
  alpha,
  power,
  alternative = c("one.sided", "two.sided"),
  type = c("golden", "nominal", "controlled")
)

```

**Arguments**

level	Threshold for the calibrated sceptical p-value. Default is 0.025.
c	Numeric vector of variance ratios of the original and replication effect estimates. This is usually the ratio of the sample size of the replication study to the sample size of the original study.

alpha	Significance level for a standard significance test in the original study. Default is 0.025.
power	Power to detect the assumed effect with a standard significance test in the original study.
alternative	Specifies if level and alpha are "two.sided" or "one.sided".
type	Type of recalibration. Can be either "golden" (default), "nominal" (no recalibration), or "controlled".

### Details

PPpSceptical is the vectorized version of the internal function .PPpSceptical\_. [Vectorize](#) is used to vectorize the function.

### Value

The project power of the sceptical p-value

### Author(s)

Leonhard Held, Samuel Pawel

### References

- Held, L. (2020). The harmonic mean chi-squared test to substantiate scientific findings. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **69**, 697-708. doi:10.1111/rssc.12410
- Held, L., Micheloud, C., Pawel, S. (2022). The assessment of replication success based on relative effect size. *The Annals of Applied Statistics*. 16:706-720. doi:10.1214/21AOAS1502
- Maca, J., Gallo, P., Branson, M., and Maurer, W. (2002). Reconsidering some aspects of the two-trials paradigm. *Journal of Biopharmaceutical Statistics*, **12**, 107-119. doi:10.1081/bip120006450

### See Also

[pSceptical](#), [levelSceptical](#), [T1EpSceptical](#)

### Examples

```
## compare project power for different recalibration types
types <- c("nominal", "golden", "controlled")
c <- seq(0.4, 5, by = 0.01)
alpha <- 0.025
power <- 0.9
pp <- sapply(X = types, FUN = function(t) {
  PPpSceptical(type = t, c = c, alpha, power, alternative = "one.sided",
    level = 0.025)
})

## compute project power of 2 trials rule
za <- qnorm(p = 1 - alpha)
mu <- za + qnorm(p = power)
```

```

pp2TR <- power * pnorm(q = za, mean = sqrt(c) * mu, lower.tail = FALSE)

matplot(x = c, y = pp * 100, type = "l", lty = 1, lwd = 2, las = 1, log = "x",
        xlab = bquote(italic(c)), ylab = "Project power (%)", xlim = c(0.4, 5),
        ylim = c(0, 100))
lines(x = c, y = pp2TR * 100, col = length(types) + 1, lwd = 2)
abline(v = 1, lty = 2)
abline(h = 90, lty = 2, col = "lightgrey")
legend("bottomright", legend = c(types, "2TR"), lty = 1, lwd = 2,
      col = seq(1, length(types) + 1))

```

---

predictionInterval      *Prediction interval for effect estimate of replication study*

---

### Description

Computes a prediction interval for the effect estimate of the replication study.

### Usage

```

predictionInterval(
  thetao,
  seo,
  ser,
  tau = 0,
  conf.level = 0.95,
  designPrior = "predictive"
)

```

### Arguments

thetao	Numeric vector of effect estimates from original studies.
seo	Numeric vector of standard errors of the original effect estimates.
ser	Numeric vector of standard errors of the replication effect estimates.
tau	Between-study heterogeneity standard error. Default is 0 (no heterogeneity). Is only taken into account when designPrior is "predictive" or "EB".
conf.level	The confidence level of the prediction intervals. Default is 0.95.
designPrior	Either "predictive" (default), "conditional", or "EB". If "EB", the contribution of the original study to the predictive distribution is shrunken towards zero based on the evidence in the original study (with empirical Bayes).

### Details

This function computes a prediction interval and a mean estimate under a specified predictive distribution of the replication effect estimate. Setting designPrior = "conditional" is not recommended since this ignores the uncertainty of the original effect estimate. See Patil, Peng, and Leek (2016) and Pawel and Held (2020) for details.

predictionInterval is the vectorized version of .predictionInterval\_. [Vectorize](#) is used to vectorize the function.

### Value

A data frame with the following columns

lower	Lower limit of prediction interval,
mean	Mean of predictive distribution,
upper	Upper limit of prediction interval.

### Author(s)

Samuel Pawel

### References

Patil, P., Peng, R. D., Leek, J. T. (2016). What should researchers expect when they replicate studies? A statistical view of replicability in psychological science. *Perspectives on Psychological Science*, **11**, 539-544. [doi:10.1177/1745691616646366](https://doi.org/10.1177/1745691616646366)

Pawel, S., Held, L. (2020). Probabilistic forecasting of replication studies. *PLoS ONE*. **15**, e0231416. [doi:10.1371/journal.pone.0231416](https://doi.org/10.1371/journal.pone.0231416)

### Examples

```
predictionInterval(theta0 = c(1.5, 2, 5), seo = 1, ser = 0.5, designPrior = "EB")

# compute prediction intervals for replication projects
data("RProjects", package = "ReplicationSuccess")
parOld <- par(mfrow = c(2, 2))
for (p in unique(RProjects$project)) {
  data_project <- subset(RProjects, project == p)
  PI <- predictionInterval(theta0 = data_project$fiso, seo = data_project$se_fiso,
                           ser = data_project$se_fisr)
  PI <- tanh(PI) # transforming back to correlation scale
  within <- (data_project$rr < PI$upper) & (data_project$rr > PI$lower)
  coverage <- mean(within)
  color <- ifelse(within == TRUE, "#333333B3", "#8B0000B3")
  study <- seq(1, nrow(data_project))
  plot(data_project$rr, study, col = color, pch = 20,
        xlim = c(-0.5, 1), xlab = expression(italic(r)[r]),
        main = paste0(p, ": ", round(coverage*100, 1), "% coverage"))
  arrows(PI$lower, study, PI$upper, study, length = 0.02, angle = 90,
         code = 3, col = color)
  abline(v = 0, lty = 3)
}
par(parOld)
```

---

pReplicate

*Probability of replicating an effect by Killeen (2005)*

---

### Description

Computes the probability that a replication study yields an effect estimate in the same direction as in the original study.

### Usage

```
pReplicate(  
  po = NULL,  
  zo = p2z(p = po, alternative = alternative),  
  c,  
  alternative = "two.sided"  
)
```

### Arguments

po	Numeric vector of p-values from the original study, default is NULL.
zo	Numeric vector of z-values from the original study. Is calculated from po, if necessary.
c	The ratio of the variances of the original and replication effect estimates. This is usually the ratio of the sample size of the replication study to the sample size of the original study.
alternative	Either "two.sided" (default) or "one.sided". Specifies whether the p-value is two-sided or one-sided.

### Details

This extends the statistic `p_rep` ("the probability of replicating an effect") by Killeen (2005) to the case of possibly unequal sample sizes, see also Senn (2002).

### Value

The probability that a replication study yields an effect estimate in the same direction as in the original study.

### Author(s)

Leonhard Held



## References

- Killeen, P. R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, **16**, 345–353. doi:10.1111/j.09567976.2005.01538.x
- Senn, S. (2002). Letter to the Editor, *Statistics in Medicine*, **21**, 2437–2444.
- Held, L. (2019). The assessment of intrinsic credibility and a new argument for  $p < 0.005$ . *Royal Society Open Science*, **6**, 181534. doi:10.1098/rsos.181534

## Examples

```
pReplicate(po = c(0.05, 0.01, 0.001), c = 1)
pReplicate(po = c(0.05, 0.01, 0.001), c = 2)
pReplicate(po = c(0.05, 0.01, 0.001), c = 2, alternative = "one.sided")
pReplicate(zo = c(2, 3, 4), c = 1)
```

---

 protzko2020

*Data from Protzko et al. (2020)*


---

## Description

Data from "High Replicability of Newly-Discovered Social-behavioral Findings is Achievable" by Protzko et al. (2020). The variables are as follows:

experiment Experiment name  
 type Type of study, either "original", "self-replication", or "external-replication"  
 lab The lab which conducted the study, either 1, 2, 3, or 4.  
 smd Standardized mean difference effect estimate  
 se Standard error of standardized mean difference effect estimate  
 n Total sample size of the study

## Usage

```
data("protzko2020")
```

## Format

A data frame with 80 rows and 6 variables

## Details

This data set originates from a prospective replication project involving four laboratories. Each of them conducted four original studies and for each original study a replication study was carried out within the same lab (self-replication) and by the other three labs (external-replication). Most studies used simple between-subject designs with two groups and a continuous outcome so that for each study, an estimate of the standardized mean difference (SMD) could be computed from the group means, group standard deviations, and group sample sizes. For studies with covariate adjustment and/or binary outcomes, effect size transformations as described in the supplementary material of

Protzko (2020) were used to obtain effect estimates and standard errors on SMD scale. The data set is licensed under a CC-BY Attribution 4.0 International license, see <https://creativecommons.org/licenses/by/4.0/> for the terms of reuse.

## Source

The relevant files were downloaded from <https://osf.io/42ef9/> on January 24, 2022. The R markdown script "Decline effects main analysis.Rmd" was executed and the relevant variables from the objects "ES\_experiments" and "decline\_effects" were saved.

## References

Protzko, J., Krosnick, J., Nelson, L. D., Nosek, B. A., Axt, J., Berent, M., ... Schooler, J. (2020, September 10). High Replicability of Newly-Discovered Social-behavioral Findings is Achievable. [doi:10.31234/osf.io/n2a9x](https://doi.org/10.31234/osf.io/n2a9x)

Protzko, J., Berent, M., Buttrick, N., DeBell, M., Roeder, S. S., Walleczek, J., ... Nosek, B. A. (2021, January 5). Results & Data. Retrieved from <https://osf.io/42ef9/>

## Examples

```
data("protzko2020", package = "ReplicationSuccess")

## forestplots of effect estimates
graphics.off()
parOld <- par(mar = c(5, 8, 4, 2), mfrow = c(4, 4))
experiments <- unique(protzko2020$experiment)
for (ex in experiments) {
  ## compute CIs
  dat <- subset(protzko2020, experiment == ex)
  za <- qnorm(p = 0.975)
  plotDF <- data.frame(lower = dat$smd - za*dat$se,
                       est = dat$smd,
                       upper = dat$smd + za*dat$se)
  colpalette <- c("#000000", "#1B9E77", "#D95F02")
  cols <- colpalette[dat$type]
  yseq <- seq(1, nrow(dat))

  ## forestplot
  plot(x = plotDF$est, y = yseq, xlim = c(-0.15, 0.8),
       ylim = c(0.8*min(yseq), 1.05*max(yseq)), type = "n",
       yaxt = "n", xlab = "Effect estimate (SMD)", ylab = "")
  abline(v = 0, col = "#0000004D")
  arrows(x0 = plotDF$lower, x1 = plotDF$upper, y0 = yseq, angle = 90,
        code = 3, length = 0.05, col = cols)
  points(y = yseq, x = plotDF$est, pch = 20, lwd = 2, col = cols)
  axis(side = 2, at = yseq, las = 1, labels = dat$type, cex.axis = 0.85)
  title(main = ex)
}
par(parOld)
```

---

pSceptical

*Computes the sceptical p-value and z-value*

---

### Description

Computes sceptical p-values and z-values based on the z-values of the original and the replication study and the corresponding variance ratios. If specified, the sceptical p-values are recalibrated.

### Usage

```
pSceptical(  
  zo,  
  zr,  
  c,  
  alternative = c("one.sided", "two.sided"),  
  type = c("golden", "nominal", "controlled")  
)  
  
zSceptical(zo, zr, c)
```

### Arguments

zo	Numeric vector of z-values from original studies.
zr	Numeric vector of z-values from replication studies.
c	Numeric vector of variance ratios of the original and replication effect estimates. This is usually the ratio of the sample size of the replication study to the sample size of the original study.
alternative	Either "one.sided" (default) or "two.sided". If "one.sided", the sceptical p-value is based on a one-sided assessment of replication success in the direction of the original effect estimate. If "two.sided", the sceptical p-value is based on a two-sided assessment of replication success regardless of the direction of the original and replication effect estimate.
type	Type of recalibration. Can be either "golden" (default), "nominal", or "controlled". Setting type to "nominal" corresponds to no recalibration as in Held et al. (2020). A recalibration is applied if type is "controlled", or "golden", and the sceptical p-value can then be interpreted on the same scale as an ordinary p-value (e.g., a one-sided sceptical p-value can be thresholded at the conventional 0.025 level). "golden" ensures that for an original study just significant at the specified level, replication success is only possible if the replication effect estimate is at least as large as the original one. "controlled" ensures exact overall Type-I error control at level $\text{level}^2$ .

### Details

pSceptical is the vectorized version of the internal function .pSceptical\_. [Vectorize](#) is used to vectorize the function.

**Value**

pSceptical returns the sceptical p-value.

zSceptical returns the z-value of the sceptical p-value.

**Author(s)**

Leonhard Held

**References**

Held, L. (2020). A new standard for the analysis and design of replication studies (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **183**, 431-448. doi:10.1111/rssa.12493

Held, L., Micheloud, C., Pawel, S. (2022). The assessment of replication success based on relative effect size. *The Annals of Applied Statistics*. 16:706-720. doi:10.1214/21AOAS1502

Micheloud, C., Balabdaoui, F., Held, L. (2023). Beyond the two-trials rule: Type-I error control and sample size planning with the sceptical p-value. <https://arxiv.org/abs/2207.00464>

**See Also**

[sampleSizeReplicationSuccess](#), [powerReplicationSuccess](#), [levelSceptical](#)

**Examples**

```
## no recalibration (type = "nominal") as in Held (2020)
pSceptical(zo = p2z(0.01), zr = p2z(0.02), c = 2, alternative = "one.sided",
           type = "nominal")

## recalibration with golden level as in Held, Micheloud, Pawel (2020)
pSceptical(zo = p2z(0.01), zr = p2z(0.02), c = 2, alternative = "one.sided",
           type = "golden")

## two-sided p-values 0.01 and 0.02, relative sample size 2
pSceptical(zo = p2z(0.01), zr = p2z(0.02), c = 2, alternative = "one.sided")
## reverse the studies
pSceptical(zo = p2z(0.02), zr = p2z(0.01), c = 1/2, alternative = "one.sided")
## both p-values 0.01, relative sample size 2
pSceptical(zo = p2z(0.01), zr = p2z(0.01), c = 2, alternative = "two.sided")

zSceptical(zo = 2, zr = 3, c = 2)
zSceptical(zo = 3, zr = 2, c = 2)
```

---

pvalueBound	<i>Bound for the p-values entering the harmonic mean chi-squared test</i>
-------------	---

---

### Description

Necessary or sufficient bounds for significance of the harmonic mean chi-squared test are computed for  $n$  one-sided p-values.

### Usage

```
pvalueBound(alpha, n, type = c("necessary", "sufficient"))
```

### Arguments

alpha	Numeric vector specifying the significance level.
n	The number of p-values.
type	Either "necessary" (default) or "sufficient". If "necessary", the necessary bounds are computed. If "sufficient", the sufficient bounds are computed.

### Value

The bound for the p-values.

### Author(s)

Leonhard Held

### References

Held, L. (2020). The harmonic mean chi-squared test to substantiate scientific findings. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **69**, 697-708. doi:10.1111/rssc.12410

### See Also

[hMeanChiSq](#)

### Examples

```
pvalueBound(alpha = 0.025^2, n = 2, type = "necessary")
pvalueBound(alpha = 0.025^2, n = 2, type = "sufficient")
```

---

Qtest	<i>Q-test to assess compatibility between original and replication effect estimate</i>
-------	--

---

**Description**

Computes p-value from meta-analytic Q-test to assess compatibility between original and replication effect estimate.

**Usage**

```
Qtest(thetao, thetar, seo, ser)
```

**Arguments**

thetao	Numeric vector of effect estimates from original studies.
thetar	Numeric vector of effect estimates from replication studies.
seo	Numeric vector of standard errors of the original effect estimates.
ser	Numeric vector of standard errors of the replication effect estimates.

**Details**

This function computes the p-value from a meta-analytic Q-test assessing compatibility between original and replication effect estimate. Rejecting compatibility when the p-value is smaller than alpha is equivalent with rejecting compatibility based on a (1 - alpha) prediction interval.

**Value**

p-value from Q-test.

**Author(s)**

Samuel Pawel

**References**

Hedges, L. V., Schauer, J. M. (2019). More Than One Replication Study Is Needed for Unambiguous Tests of Replication. *Journal of Educational and Behavioral Statistics*, **44**, 543-570. [doi:10.3102/1076998619852953](https://doi.org/10.3102/1076998619852953)

**See Also**

[predictionInterval](#)

**Examples**

```
Qtest(thetao = 2, thetar = 0.5, seo = 1, ser = 0.5)
```

RProjects

*Data from four large-scale replication projects***Description**

Data from *Reproducibility Project Psychology (RPP)*, *Experimental Economics Replication Project (EERP)*, *Social Sciences Replication Project (SSRP)*, *Experimental Philosophy Replicability Project (EPRP)*. The variables are as follows:

study Study identifier, usually names of authors from original study  
 project Name of replication project  
 ro Effect estimate of original study on correlation scale  
 rr Effect estimate of replication study on correlation scale  
 fiso Effect estimate of original study transformed to Fisher-z scale  
 fisr Effect estimate of replication study transformed to Fisher-z scale  
 se\_fiso Standard error of Fisher-z transformed effect estimate of original study  
 se\_fisr Standard error of Fisher-z transformed effect estimate of replication study  
 po Two-sided p-value from significance test of effect estimate from original study  
 pr Two-sided p-value from significance test of effect estimate from replication study  
 po1 One-sided p-value from significance test of effect estimate from original study (in the direction of the original effect estimate)  
 pr1 One-sided p-value from significance test of effect estimate from replication study (in the direction of the original effect estimate)  
 pm\_belief Peer belief about whether replication effect estimate will achieve statistical significance elicited through prediction market (only available for EERP and SSRP)  
 no Sample size in original study  
 nr Sample size in replication study

**Usage**

```
data(RProjects)
```

**Format**

A data frame with 143 rows and 15 variables

**Details**

Two-sided p-values were calculated assuming normality of Fisher-z transformed effect estimates. From the RPP only the *meta-analytic subset* is included, which consists of 73 out of 100 study pairs for which the standard error of the z-transformed correlation coefficient can be computed. For the RPP sample sizes were recalculated from the reported standard errors of Fisher z-transformed correlation coefficients. From the EPRP only 31 out of 40 study pairs are included where effective sample size for original and replication study are available simultaneously. For more details about how the the data was preprocessed see source below and supplement S1 of Pawel and Held (2020).

## Source

RPP: The source files were downloaded from <https://github.com/CenterForOpenScience/rpp/>. The "masterscript.R" file was executed and the relevant variables were extracted from the generated "final" object (standard errors of Fisher-z transformed correlations) and "MASTER" object (everything else). The data set is licensed under a CC0 1.0 Universal license, see <https://creativecommons.org/publicdomain/zero/1.0/> for the terms of reuse.

EERP: The source files were downloaded from <https://osf.io/pnwuz/>. The required data were then manually extracted from the code in the files "effectdata.py" (sample sizes) and "create\_studydetails.do" (everything else). Data regarding the prediction market and survey beliefs were manually extracted from table S3 of the supplementary materials of the EERP. The authors of this R package have been granted permission to share this data set by the coordinators of the EERP.

SSRP: The relevant variables were extracted from the file "D3 - ReplicationResults.csv" downloaded from <https://osf.io/abu7k>. For replications which underwent only the first stage, the data from the first stage were taken as the data for the replication study. For the replications which reached the second stage, the pooled data from both stages were taken as the data for the replication study. Data regarding survey and prediction market beliefs were extracted from the "D6 - MeanPeerBeliefs.csv" file, which was downloaded from <https://osf.io/vr6p8/>. The data set is licensed under a CC0 1.0 Universal license, see <https://creativecommons.org/publicdomain/zero/1.0/> for the terms of reuse.

EPRP: Data were taken from the "XPhiReplicability\_CompleteData.csv" file, which was downloaded from <https://osf.io/4ewkh/>. The authors of this R package have been granted permission to share this data set by the coordinators of the EPRP.

## References

- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., ... Hang, W. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, **351**, 1433-1436. doi:10.1126/science.aaf0918
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., ... Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, **2**, 637-644. doi:10.1038/s415620180399z
- Cova, F., Strickland, B., Abatista, A., Allard, A., Andow, J., Attie, M., ... Zhou, X. (2018). Estimating the reproducibility of experimental philosophy. *Review of Philosophy and Psychology*. doi:10.1007/s1316401804009
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, **349**, aac4716. doi:10.1126/science.aac4716
- Pawel, S., Held, L. (2020). Probabilistic forecasting of replication studies. *PLoS ONE*. **15**, e0231416. doi:10.1371/journal.pone.0231416

## See Also

[SSRP](#)

## Examples

```
data("RProjects", package = "ReplicationSuccess")
```



```

## Computing key quantities
RProjects$zo <- RProjects$fico/RProjects$se_fiso
RProjects$zr <- RProjects$fir/RProjects$se_fisr
RProjects$c <- RProjects$se_fiso^2/RProjects$se_fisr^2

## Computing one-sided p-values for alternative = "greater"
RProjects$po1 <- z2p(z = RProjects$zo, alternative = "greater")
RProjects$pr1 <- z2p(z = RProjects$zr, alternative = "greater")

## Plots of effect estimates
parOld <- par(mfrow = c(2, 2))
for (p in unique(RProjects$project)) {
  data_project <- subset(RProjects, project == p)
  plot(rr ~ ro, data = data_project, ylim = c(-0.5, 1),
       xlim = c(-0.5, 1), main = p, xlab = expression(italic(r)[o]),
       ylab = expression(italic(r)[r]))
  abline(h = 0, lty = 2)
  abline(a = 0, b = 1, col = "grey")
}
par(parOld)

## Plots of peer beliefs
RProjects$significant <- factor(RProjects$pr < 0.05,
                              levels = c(FALSE, TRUE),
                              labels = c("no", "yes"))
parOld <- par(mfrow = c(1, 2))
for (p in c("Experimental Economics", "Social Sciences")) {
  data_project <- subset(RProjects, project == p)
  boxplot(pm_belief ~ significant, data = data_project, ylim = c(0, 1),
         main = p, xlab = "Replication effect significant", ylab = "Peer belief")
  stripchart(pm_belief ~ significant, data = data_project, vertical = TRUE,
            add = TRUE, pch = 1, method = "jitter")
}
par(parOld)

## Computing the sceptical p-value
ps <- with(RProjects, pSceptical(zo = fico/se_fiso,
                                zr = fir/se_fisr,
                                c = se_fiso^2/se_fisr^2))

```

---

```
sampleSizeReplicationSuccess
```

*Computes the required relative sample size to achieve replication success with the sceptical p-value*

---

## Description

The relative sample size to achieve replication success is computed based on the z-value of the original study, the type of recalibration, the power and the design prior.

**Usage**

```
sampleSizeReplicationSuccess(
  zo,
  power = NA,
  level = 0.025,
  alternative = c("one.sided", "two.sided"),
  type = c("golden", "nominal", "controlled"),
  designPrior = c("conditional", "predictive", "EB"),
  shrinkage = 0,
  h = 0
)
```

**Arguments**

zo	Numeric vector of z-values from original studies.
power	The power to achieve replication success.
level	Threshold for the calibrated sceptical p-value. Default is 0.025.
alternative	Specifies if level is "one.sided" (default) or "two.sided". If "one.sided" then sample size calculations are based on a one-sided assessment of replication success in the direction of the original effect estimates.
type	Type of recalibration. Can be either "golden" (default), "nominal" (no recalibration), or "controlled". "golden" ensures that for an original study just significant at the specified level, replication success is only possible for replication effect estimates larger than the original one. "controlled" ensures exact overall Type-I error control at level $\text{level}^2$ .
designPrior	Is only taken into account when power is specified. Either "conditional" (default), "predictive", or "EB". If "EB", the power is computed under a predictive distribution where the contribution of the original study is shrunken towards zero based on the evidence in the original study (with an empirical Bayes shrinkage estimator).
shrinkage	Is only taken into account when power is specified. A number in [0,1) with default 0. Specifies the shrinkage of the original effect estimate towards zero (e.g., the effect is shrunken by a factor of 25% for shrinkage = 0.25). Is only taken into account when the designPrior is "conditional" or "predictive".
h	Is only taken into account when power is specified and designPrior is "predictive" or "EB". The relative between-study heterogeneity, i.e., the ratio of the heterogeneity variance to the variance of the original effect estimate. Default is 0 (no heterogeneity).

**Details**

sampleSizeReplicationSuccess is the vectorized version of the internal function `.sampleSizeReplicationSuccess_`. [Vectorize](#) is used to vectorize the function.

**Value**

The relative sample size for replication success. If impossible to achieve the desired power for specified inputs NaN is returned.

**Author(s)**

Leonhard Held, Charlotte Micheloud, Samuel Pawel, Florian Gerber

**References**

- Held, L. (2020). A new standard for the analysis and design of replication studies (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **183**, 431-448. doi:10.1111/rssa.12493
- Held, L., Micheloud, C., Pawel, S. (2022). The assessment of replication success based on relative effect size. *The Annals of Applied Statistics*. 16:706-720. doi:10.1214/21AOAS1502
- Micheloud, C., Balabdaoui, F., Held, L. (2023). Beyond the two-trials rule: Type-I error control and sample size planning with the sceptical p-value. <https://arxiv.org/abs/2207.00464>

**See Also**

[pSceptical](#), [powerReplicationSuccess](#), [levelSceptical](#)

**Examples**

```
## based on power
sampleSizeReplicationSuccess(zo = p2z(0.0025), power = 0.8, level = 0.025,
                             type = "golden")
sampleSizeReplicationSuccess(zo = p2z(0.0025), power = 0.8, level = 0.025,
                             type = "golden", designPrior = "predictive")
```

---

sampleSizeSignificance

*Computes the required relative sample size to achieve significance based on power*

---

**Description**

The relative sample size to achieve significance of the replication study is computed based on the z-value of the original study, the significance level and the power.

**Usage**

```
sampleSizeSignificance(
  zo,
  power = NA,
  level = 0.025,
  alternative = c("one.sided", "two.sided"),
```

```

designPrior = c("conditional", "predictive", "EB"),
h = 0,
shrinkage = 0
)

```

### Arguments

zo	A vector of z-values from original studies.
power	The power to achieve replication success.
level	Significance level. Default is 0.025.
alternative	Either "one.sided" (default) or "two.sided". Specifies if the significance level is one-sided or two-sided. If the significance level is one-sided, then sample size calculations are based on a one-sided assessment of significance in the direction of the original effect estimate.
designPrior	Is only taken into account when power is specified. Either "conditional" (default), "predictive", or "EB". If "EB", the power is computed under a predictive distribution where the contribution of the original study is shrunken towards zero based on the evidence in the original study (with an empirical Bayes shrinkage estimator).
h	Is only taken into account when power is specified and designPrior is "predictive" or "EB". The relative between-study heterogeneity, i.e., the ratio of the heterogeneity variance to the variance of the original effect estimate. Default is 0 (no heterogeneity).
shrinkage	Is only taken into account when power is specified. A number in [0,1) with default 0. Specifies the shrinkage of the original effect towards zero (e.g., shrinkage = 0.25 implies shrinkage by a factor of 25%). Is only taken into account when designPrior is "conditional" or "predictive".

### Details

sampleSizeSignificance is the vectorized version of .sampleSizeSignificance\_. [Vectorize](#) is used to vectorize the function.

### Value

The relative sample size to achieve significance in the specified direction. If impossible to achieve the desired power for specified inputs NaN is returned.

### Author(s)

Leonhard Held, Samuel Pawel, Charlotte Micheloud, Florian Gerber

### References

Held, L. (2020). A new standard for the analysis and design of replication studies (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **183**, 431-448. doi:10.1111/rssa.12493

Pawel, S., Held, L. (2020). Probabilistic forecasting of replication studies. *PLoS ONE*. **15**, e0231416. doi:10.1371/journal.pone.0231416

Held, L., Micheloud, C., Pawel, S. (2022). The assessment of replication success based on relative effect size. *The Annals of Applied Statistics*. 16:706-720. doi:10.1214/21AOAS1502

Micheloud, C., Held, L. (2022). Power Calculations for Replication Studies. *Statistical Science*. 37:369-379. doi:10.1214/21STS828

### See Also

[powerSignificance](#)

### Examples

```
sampleSizeSignificance(zo = p2z(0.005), power = 0.8)
sampleSizeSignificance(zo = p2z(0.005, alternative = "two.sided"), power = 0.8)
sampleSizeSignificance(zo = p2z(0.005), power = 0.8, designPrior = "predictive")

sampleSizeSignificance(zo = 3, power = 0.8, designPrior = "predictive",
                      shrinkage = 0.5, h = 0.25)
sampleSizeSignificance(zo = 3, power = 0.8, designPrior = "EB", h = 0.5)

# sample size to achieve 0.8 power as function of original p-value
zo <- p2z(seq(0.0001, 0.05, 0.0001))
oldPar <- par(mfrow = c(1,2))
plot(z2p(zo), sampleSizeSignificance(zo = zo, designPrior = "conditional", power = 0.8),
     type = "l", ylim = c(0.5, 10), log = "y", lwd = 1.5, ylab = "Relative sample size",
     xlab = expression(italic(p)[o]), las = 1)
lines(z2p(zo), sampleSizeSignificance(zo = zo, designPrior = "predictive", power = 0.8),
     lwd = 2, lty = 2)
lines(z2p(zo), sampleSizeSignificance(zo = zo, designPrior = "EB", power = 0.8),
     lwd = 1.5, lty = 3)
legend("topleft", legend = c("conditional", "predictive", "EB"),
     title = "Design prior", lty = c(1, 2, 3), lwd = 1.5, bty = "n")

par(oldPar)
```

### Description

Data from the *Social Sciences Replication Project* (SSRP) including the details of the interim analysis. The variables are as follows:

**study** Study identifier, usually names of authors from original study

**ro** Effect estimate of original study on correlation scale

**ri** Effect estimate of replication study at the interim analysis on correlation scale

**rr** Effect estimate of replication study at the final analysis on correlation scale

fiso Effect estimate of original study transformed to Fisher-z scale  
 fisi Effect estimate of replication study at the interim analysis transformed to Fisher-z scale  
 fisr Effect estimate of replication study at the final analysis transformed to Fisher-z scale  
 se\_fiso Standard error of Fisher-z transformed effect estimate of original study  
 se\_fisi Standard error of Fisher-z transformed effect estimate of replication study at the interim analysis  
 se\_fisr Standard error of Fisher-z transformed effect estimate of replication study at the final analysis  
 no Sample size in original study  
 ni Sample size in replication study at the interim analysis  
 nr Sample size in replication study at the final analysis  
 po Two-sided p-value from significance test of effect estimate from original study  
 pi Two-sided p-value from significance test of effect estimate from replication study at the interim analysis  
 pr Two-sided p-value from significance test of effect estimate from replication study at the final analysis  
 n75 Sample size calculated to have 90% power in replication study to detect 75% of the original effect size (expressed as the correlation coefficient r)  
 n50 Sample size calculated to have 90% power in replication study to detect 50% of the original effect size (expressed as the correlation coefficient r)

### Usage

data(SSRP)

### Format

A data frame with 21 rows and 18 variables

### Details

Two-sided p-values were calculated assuming normality of Fisher-z transformed effect estimates. A two-stage procedure was used for the replications. In stage 1, the authors had 90% power to detect 75% of the original effect size at the 5% significance level in a two-sided test. If the original result replicated in stage 1 (two-sided P-value < 0.05 and effect in the same direction as in the original study), the data collection was stopped. If not, a second data collection was carried out in stage 2 to have 90% power to detect 50% of the original effect size for the first and the second data collections pooled. n75 and n50 are the planned sample sizes calculated to reach 90% power in stage 1 and 2, respectively. They sometimes differ from the sample sizes that were actually collected (ni and nr, respectively). See supplementary information of Camerer et al. (2018) for details.

### Source

<https://osf.io/abu7k>

## References

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., ... Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, **2**, 637-644. doi:10.1038/s415620180399z

## See Also

[RProjects](#)

## Examples

```
# plot of the sample sizes
plot(ni ~ no, data = SSRP, ylim = c(0, 2500), xlim = c(0, 400),
     xlab = expression(n[o]), ylab = expression(n[i]))
abline(a = 0, b = 1, col = "grey")

plot(nr ~ no, data = SSRP, ylim = c(0, 2500), xlim = c(0, 400),
     xlab = expression(n[o]), ylab = expression(n[r]))
abline(a = 0, b = 1, col = "grey")
```

---

T1EpSceptical

*Compute overall type-I error rate of the sceptical p-value*

---

## Description

The overall type-I error rate of the sceptical p-value is computed for a specified level, the relative variance, and the alternative hypothesis.

## Usage

```
T1EpSceptical(
  level,
  c,
  alternative = c("one.sided", "two.sided"),
  type = c("golden", "nominal", "controlled")
)
```

## Arguments

level	Threshold for the calibrated sceptical p-value. Default is 0.025.
c	Numeric vector of variance ratios of the original and replication effect estimates. This is usually the ratio of the sample size of the replication study to the sample size of the original study.
alternative	Specifies if level is "two.sided" or "one.sided".
type	Type of recalibration. Recalibration type can be either "golden" (default), "nominal" (no recalibration), or "controlled".

**Details**

T1EpSceptical is the vectorized version of the internal function `.T1EpSceptical_`. [Vectorize](#) is used to vectorize the function.

**Value**

The overall type-I error rate.

**Author(s)**

Leonhard Held, Samuel Pawel

**References**

Held, L. (2020). The harmonic mean chi-squared test to substantiate scientific findings. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **69**, 697-708. doi:10.1111/rssc.12410

Held, L., Micheloud, C., Pawel, S. (2022). The assessment of replication success based on relative effect size. *The Annals of Applied Statistics*. 16:706-720. doi:10.1214/21AOAS1502

Micheloud, C., Balabdaoui, F., Held, L. (2023). Beyond the two-trials rule: Type-I error control and sample size planning with the sceptical p-value. <https://arxiv.org/abs/2207.00464>

**See Also**

[pSceptical](#), [levelSceptical](#), [PPpSceptical](#)

**Examples**

```
## compare type-I error rate for different recalibration types
types <- c("nominal", "golden", "controlled")
c <- seq(0.2, 5, by = 0.05)
t1 <- sapply(X = types, FUN = function(t) {
  T1EpSceptical(type = t, c = c, alternative = "one.sided", level = 0.025)
})
matplot(x = c, y = t1*100, type = "l", lty = 1, lwd = 2, las = 1, log = "x",
        xlab = bquote(italic(c)), ylab = "Type-I error (%)", xlim = c(0.2, 5))
legend("topright", legend = types, lty = 1, lwd = 2, col = seq_along(types))
```

---

thresholdIntrinsic      *Computes the p-value threshold for intrinsic credibility*

---

**Description**

Computes the p-value threshold for intrinsic credibility



**Usage**

```
thresholdIntrinsic(  
  alpha,  
  alternative = c("two.sided", "one.sided"),  
  type = c("Held", "Matthews")  
)
```

**Arguments**

alpha	Numeric vector of intrinsic credibility levels.
alternative	Either "two.sided" (default) or "one.sided". Specifies if the threshold is for one-sided or two-sided p-values.
type	Either "Held" (default) or "Matthews". Type of intrinsic p-value threshold, see Held (2019) and Matthews (2018) for more information.

**Value**

The threshold for intrinsic credibility.

**Author(s)**

Leonhard Held

**References**

Matthews, R. A. J. (2018). Beyond 'significance': principles and practice of the analysis of credibility. *Royal Society Open Science*, **5**, 171047. doi:[10.1098/rsos.171047](https://doi.org/10.1098/rsos.171047)

Held, L. (2019). The assessment of intrinsic credibility and a new argument for  $p < 0.005$ . *Royal Society Open Science*, **6**, 181534. doi:[10.1098/rsos.181534](https://doi.org/10.1098/rsos.181534)

**Examples**

```
thresholdIntrinsic(alpha = c(0.005, 0.01, 0.05))  
thresholdIntrinsic(alpha = c(0.005, 0.01, 0.05), alternative = "one.sided")
```

# Index

## \* data

protzko2020, 25  
RProjects, 31  
SSRP, 37

ci2estimate (ci2se), 2  
ci2p (ci2se), 2  
ci2se, 2  
ci2z (ci2se), 2

effectSizeReplicationSuccess, 4, 7  
effectSizeSignificance, 6

hMeanChiSq, 7, 29  
hMeanChiSqCI (hMeanChiSq), 7  
hMeanChiSqMu (hMeanChiSq), 7

levelSceptical, 5, 10, 15, 21, 28, 35, 40

p2z (ci2se), 2  
pBox, 11  
pIntrinsic, 13  
powerReplicationSuccess, 14, 28, 35  
powerSignificance, 16, 19, 37  
powerSignificanceInterim, 17, 18  
PPpSceptical, 20, 40  
predictionInterval, 22, 30  
pReplicate, 24  
protzko2020, 25  
pSceptical, 15, 21, 27, 35, 40  
pvalueBound, 29

Qtest, 30

RProjects, 31, 39

sampleSizeReplicationSuccess, 5, 15, 28,  
33  
sampleSizeSignificance, 17, 19, 35  
SSRP, 32, 37

T1EpSceptical, 21, 39

thresholdIntrinsic, 40

Vectorize, 3, 5, 6, 11, 15, 17, 19, 21, 23, 27,  
34, 36, 40

z2p (ci2se), 2  
zBox (pBox), 11  
zSceptical (pSceptical), 27