

Package: BUCSS (via r-universe)

October 13, 2026

Type Package

Title Bias and Uncertainty Corrected Sample Size

Version 2.0.0

Description Implements sample size planning for a future study from the summary statistics of a prior study, correcting the prior effect for publication bias and uncertainty so the planned study attains its target statistical power more reliably than methods that take a sample effect size at face value. Covers the independent and dependent 't' test, between-subjects, within-subjects, and split-plot factorial analysis of variance, and multiple regression. The method is described in Anderson, Kelley, and Maxwell (2017) <doi:10.1177/0956797617723724>. The noncentrality parameter underlying the correction is obtained by direct numerical root finding, a more accurate approach than a fixed grid search that also removes an internal upper limit on the parameter. A small number of planned sample sizes therefore differ slightly from versions 1.x; the new values are more accurate.

License GPL (>= 3)

Encoding UTF-8

Language en-US

Depends R (>= 3.5.0)

Imports generics, graphics, stats

Suggests testthat (>= 3.0.0), knitr, rmarkdown

VignetteBuilder knitr

URL <https://github.com/yelleKneK/BUCSS>,
<https://yelleKneK.github.io/BUCSS>,
<https://designingexperiments.com>

BugReports <https://github.com/yelleKneK/BUCSS/issues>

Config/testthat/edition 3

RoxygenNote 7.3.2
Repository <https://yelleknek.r-universe.dev>
Date/Publication 2026-07-30 01:29:47 UTC
RemoteUrl <https://github.com/yelleknek/bucss>
RemoteRef HEAD
RemoteSha 4c511c43bc81c8b2638569d4b3540408da36b550

Contents

BUCSS-package	2
bucss_tidiers	3
planning_sentence	5
plot.bucss_power	5
ss_buc_ancova	8
ss_buc_chisq_diff	11
ss_buc_correlation	15
ss_buc_factorial_anova	19
ss_buc_factorial_anova_general	22
ss_buc_independent_t	26
ss_buc_manova	30
ss_buc_mixed_anova	34
ss_buc_mixed_anova_general	37
ss_buc_one_sample_t	41
ss_buc_one_way_anova	43
ss_buc_paired_t	46
ss_buc_R2	50
ss_buc_reg_coef	53
ss_buc_reg_joint	56
ss_buc_rm_anova	60
ss_buc_rm_anova_general	63
ss_buc_sensitivity	66
ss_buc_welch_t	68
Index	73

BUCSS-package	<i>BUCSS: Bias and Uncertainty Corrected Sample Size</i>
---------------	--

Description

Implements sample size planning for a future study from the summary statistics of a prior study, correcting the prior effect for publication bias and uncertainty so the planned study attains its target statistical power more reliably than methods that take a sample effect size at face value. The correction builds on the truncated noncentral distribution of Taylor and Muller (1996) and is described in Anderson, Kelley, and Maxwell (2017).

Details

The functions are also available, through a web interface, as Shiny apps at <https://designingexperiments.com>.

BUCSS follows the design and naming conventions of the **DMAR** package (Design, Measurement, and Analysis of Human-Centered Research): its `ss_buc_*` planners parallel the `ss_aipe_*` (accuracy in parameter estimation) and `ss_power_*` (power analysis) sample size planning families in **DMAR**, returning a tidy result object in the same house style. **DMAR** is a natural companion for effect size estimation, confidence intervals, and sample size planning across additional frameworks.

The bias and uncertainty adjusted noncentrality parameter is found by solving the truncated-likelihood equation directly with `uniroot`, so it is no longer capped at 100 (the 1.x grid bound) and needs no resolution argument.

Author(s)

Ken Kelley (<kkelley@nd.edu>), Samantha F. Anderson (<samantha.f.anderson@asu.edu>), and Scott E. Maxwell

References

Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:[10.1037/met0000520](https://doi.org/10.1037/met0000520)

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:[10.1177/0956797617723724](https://doi.org/10.1177/0956797617723724)

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.

See Also

DMAR

bucss_tidiers	<i>Tidy and summarize a bias and uncertainty corrected sample size result</i>
---------------	---

Description

`tidy()` returns the compact estimate view, mirroring the sibling package **DMAR**’s sample size planners: one row with term (`"sample_size"`), estimate (the necessary sample size in the design’s unit), and power (the conservative achieved power at that sample size; see the planner’s help page). `glance()` returns the one-row wide view: the estimate plus one column per row of the result (`total_N` where applicable, `ncp_adjusted`, and every echoed planning input) together with the design, the effect tested, the unit, the planned test’s degrees of freedom, and the assurance ceiling.

Usage

```
## S3 method for class 'bucss_power'  
tidy(x, ...)  
  
## S3 method for class 'bucss_power'  
glance(x, ...)
```

Arguments

x	An object of class <code>bucss_power</code> returned by an <code>ss_buc_*</code> function.
...	Ignored.

Details

These methods come from the **generics** package and need no extra setup; the verbs are re-exported by **BUCSS**, so `tidy(result)` and `glance(result)` work as soon as the package is loaded. The result is also an ordinary `data.frame`, so the long view and the stored full precision values remain available directly, for example `result$value[result$term == "ncp_adjusted"]`.

Value

A one-row `data.frame`. For `tidy()`, the columns `term`, `estimate`, and `power`. For `glance()`, the wide summary described above.

Author(s)

Ken Kelley (<kkelley@end.edu>)

Examples

```
result <- ss_buc_factorial_anova(F_observed = 5, N = 120, levels_A = 2,  
  levels_B = 3, effect = "factor_B")  
  
# Compact estimate view, as in DMAR: the sample size and its power.  
tidy(result)  
tidy(result)$estimate  
  
# One-row wide view: every quantity and echoed input as a column.  
glance(result)  
glance(result)$total_N
```

planning_sentence	<i>The planning sentence for a manuscript</i>
-------------------	---

Description

Turns a `bucss_power` result into the sentence an author writes in a manuscript's planning or method section: the necessary sample size in its design's unit, the desired power, the assurance, and a clause noting the publication bias and uncertainty correction. This mirrors the `results_sentence()` helper in the sibling package `DMAR`.

Usage

```
planning_sentence(x)
```

Arguments

`x` An object of class `bucss_power`, as returned by any `ss_buc_*` function.

Value

A single character string.

Author(s)

Ken Kelley (<kkelley@end.edu>)

Examples

```
result <- ss_buc_independent_t(t_observed = 3, n = c(50, 55),
  assurance = .90)
planning_sentence(result)
```

plot.bucss_power	<i>Plot the assurance ceiling of a plan</i>
------------------	---

Description

Re-runs the planner that produced `x` across a range of assurance values and draws what happens as the requested assurance approaches the largest value the prior result can support: the bias and uncertainty adjusted noncentrality parameter falling to zero, and the sample size it implies growing asymptotically.

Usage

```
## S3 method for class 'bucss_power'
plot(
  x,
  which = c("both", "ncp", "size"),
  assurance = NULL,
  lower = NULL,
  n_points = 60L,
  col = "#B03A2E",
  ...
)
```

Arguments

x	A plan returned by any ss_buc_* function.
which	Which panel to draw: "both" (the default, side by side), "ncp" for the corrected noncentrality parameter alone, or "size" for the sample size alone.
assurance	Assurance values to evaluate. NULL (the default) builds a grid running from lower up to just short of the ceiling.
lower	Smallest assurance to evaluate, used only when assurance is NULL. Defaults to .5, the recommended floor, or to half the ceiling when the ceiling itself is below .5.
n_points	Number of assurance values in the automatic grid.
col	Color for the curve and the marked plan.
...	Further arguments passed to the underlying plot calls.

Details

Every prior result has a largest supportable assurance, the closed form $1 - p/\alpha$ where p is the prior study's p value and α is `alpha_prior`. Request more than that and the corrected noncentrality parameter is driven to zero, no sample size attains the target power, and the planner stops. The ceiling is reported in the printed footer of every plan and named in that error, but its shape is the point: assurance does not degrade gracefully as it rises, it hits a wall.

Two panels are drawn, sharing an assurance axis.

- **The corrected noncentrality parameter**, which falls to zero at the ceiling. This is the quantity the correction actually solves for, and the panel shows how little of it survives near the ceiling.
- **The necessary sample size**, on a logarithmic scale, because it spans orders of magnitude over the plotted range. This is the same story in the units a researcher budgets in.

The ceiling is drawn as a dashed vertical line in both panels and the plan's own assurance is marked with a filled point, so a reader can see at a glance how much headroom their prior result has left.

The grid is spaced evenly in the logarithm of the remaining headroom rather than in assurance, so that roughly half the plotted points fall in the last few hundredths below the ceiling. Spacing evenly

in assurance would draw the wall as a single jump between two grid points and lose the shape entirely.

The plotted values are returned invisibly as an ordinary data.frame, one row per assurance, so the figure can be redrawn with any other graphics system. BUCSS draws it in base graphics and takes no plotting dependency of its own.

Value

Invisibly, a data.frame with columns assurance, ncp_adjusted, sample_size, and actual_power, one row per evaluated assurance. The largest supportable assurance travels on the assurance_ceiling attribute, and the design and sample size unit on design and sample_size_unit.

Author(s)

Ken Kelley (<kkelley@end.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.
- Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

See Also

[ss_buc_sensitivity](#) to audit a plan by simulation, and vignette("understanding-assurance", package = "BUCSS") for what the ceiling means.

Examples

```
plan <- ss_buc_independent_t(t_observed = 3, n = 20, assurance = .80)

# The ceiling, drawn: the corrected parameter falls to zero and the
# sample size grows asymptotically.
plot(plan)

# The plotted values, for redrawing elsewhere.
curve_values <- plot(plan, which = "size", n_points = 20)
head(curve_values)
attr(curve_values, "assurance_ceiling")
```

ss_buc_ancova	<i>Necessary sample size to reach desired power for an analysis of covariance using a publication bias and uncertainty correction procedure</i>
---------------	---

Description

ss_buc_ancova returns the necessary per-cell sample size to achieve a desired level of statistical power for a planned study testing a group effect (omnibus or contrast) while adjusting for one or more covariates, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

Usage

```
ss_buc_ancova(
  F_observed,
  N,
  cells,
  n_covariates,
  df_numerator = cells - 1,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

F_observed	Observed F value for the effect of interest from a previous study used to plan sample size for a planned study. For a single-degree-of-freedom contrast, this is the square of the contrast's t .
N	Total sample size of the previous study.
cells	Number of cells (groups) in the between-subjects design.
n_covariates	Number of covariates adjusted for in the analysis.
df_numerator	Numerator degrees of freedom for the effect of interest. Defaults to cells - 1, the omnibus group effect; use 1 for a single-degree-of-freedom contrast.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.

assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

An analysis of covariance is a between-subjects analysis with additional estimated parameters. `ss_buc_ancova` takes the number of covariates directly and carries them through to the planned study: the prior study's error degrees of freedom are N minus the number of cells minus the number of covariates, and the planned study's error degrees of freedom are reduced by the same number of covariates. Supplying `n_covariates = 0` reproduces the corresponding [ss_buc_one_way_anova](#) or [ss_buc_factorial_anova_general](#) result exactly.

By default the omnibus group effect is planned (`df_numerator = cells - 1`). For a single-degree-of-freedom contrast among the adjusted means, supply the contrast's F (the square of its t) and set `df_numerator = 1`.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. Thus, the ratio of the numerator and the denominator is a truncated noncentral F distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, `assurance = .5` means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting `assurance = .5` (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, `assurance > .5` can be selected, which corresponds to selecting the noncentrality parameter associated with the $(1 - \text{assurance})$ quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior = .20` would reflect less severe publication bias than the default of .05. In essence, setting `alpha_prior` at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect

cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior` = 1 and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

The covariates are treated as fixed, which is the usual convention in sample size planning for an analysis of covariance. When the covariates are themselves sampled, the planned study's power is very slightly lower than reported.

`ss_buc_ancova` assumes that the planned study will have equal n . If the prior N is not equally divisible by the number of cells, the per-cell size is rounded both up and down, the planned sample size is computed under each, and the larger of the two is returned, with the smaller of the two adjusted noncentrality parameters, to be conservative.

Value

An object of class `bucss_power`: a `data.frame` with a character `term` column and a numeric `value` column, in the same shape as the sibling package `DMAR`'s planners. The design results come first: the necessary per-cell sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724

Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.

Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

See Also

[ss_buc_factorial_anova_general](#), which this function calls, and which takes the error degrees of freedom directly.

Examples

```
# Three groups, two covariates, prior N = 120: the omnibus group effect.
result <- ss_buc_ancova(F_observed = 6, N = 120, cells = 3,
  n_covariates = 2, alpha_prior = .05, alpha_planned = .05,
  assurance = .80, desired_power = .80)
result

# A single-degree-of-freedom contrast among the adjusted means.
ss_buc_ancova(F_observed = 9, N = 120, cells = 3, n_covariates = 2,
  df_numerator = 1)

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .93):
try(ss_buc_ancova(F_observed = 6, N = 120, cells = 3, n_covariates = 2,
  assurance = .95))
```

ss_buc_chisq_diff	<i>Necessary sample size to reach desired power for a nested model chi-square difference test using a publication bias and uncertainty correction procedure</i>
-------------------	---

Description

ss_buc_chisq_diff returns the necessary total sample size to achieve a desired level of statistical power for a planned study whose effect of interest is tested by the chi-square difference between two nested models (the likelihood ratio test used, for example, to test a constrained path in a structural equation model), based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned.

Usage

```
ss_buc_chisq_diff(
  chisq_observed,
  N,
  df_full,
  df_restricted,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

chisq_observed	Observed chi-square difference between the two nested models in the previous study.
N	Total sample size of the previous study.
df_full	Degrees of freedom of the full (less constrained) model. A saturated full model has zero, which identifies the comparison as an omnibus test of model fit rather than a difference test between two substantive models; the function refuses that case (see Details).
df_restricted	Degrees of freedom of the restricted (more constrained) model. It exceeds df_full by the number of constraints imposed, which is the difference test's degrees of freedom.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

Scope, and one thing this function is not for. The correction applies to a test whose statistic is large when the effect is present, so that publication selects large values. A chi-square difference

test between nested models is such a test: under the alternative it is approximately noncentral chi-square with degrees of freedom equal to the number of constraints released and a noncentrality parameter proportional to sample size, and a paper reports the constrained path as supported when the difference test is significant. The correction is therefore built for the difference test.

It is *not* appropriate for an omnibus model fit chi-square, where a publishable result is a *small* statistic (a model that fits). The selection region is inverted there, so the truncated likelihood this function builds does not describe it.

The function refuses that case rather than only warning about it. This is why both models' degrees of freedom are required rather than only their difference. An omnibus fit chi-square is the comparison of a model against the *saturated* model, and a saturated model has zero degrees of freedom, so `df_full = 0` identifies the misuse exactly. Given only the difference, nothing distinguishes a fit test from a genuine comparison of two substantive models, and the mistake would pass silently. Supplying both is also a more faithful description of what was tested, and the two values are what a paper reports for each model.

The correction uses a likelihood function of a truncated noncentral chi-square distribution, the same construction the rest of the package uses for the noncentral *F* (see Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017): the truncated area between the critical value and the observed statistic, divided by the power of the test. Because the noncentrality parameter of a likelihood ratio test is proportional to sample size, the planned study's noncentrality parameter is the corrected one scaled by the ratio of the planned to the prior sample size.

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, `assurance = .5` means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting `assurance = .5` (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, `assurance > .5` can be selected, which corresponds to selecting the noncentrality parameter associated with the $(1 - \text{assurance})$ quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior = .20` would reflect less severe publication bias than the default of .05. In essence, setting `alpha_prior` at .20 assumes that studies with *p* values less than .20 are published, whereas those with larger *p* values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where *p* is the prior study's *p* value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by

setting `alpha_prior = 1` and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

The chi-square difference test is asymptotic, so both the correction and the planned-study power inherit that approximation. The test is liberal in small samples: in simulation its actual Type I error is nearer .06 than .05 at a total sample size of 50, and it reaches the nominal rate only around 200. Treat a planned sample size under a hundred or so with caution, and read a very small one as a signal that the prior effect was large rather than as a serious recommendation.

Value

An object of class `bucss_power`: a `data.frame` with a character `term` column and a numeric `value` column, in the same shape as the sibling package `DMAR`'s planners. The design results come first: the necessary total sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.
- Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

Examples

```
# A path constrained to zero: the full model has 42 degrees of freedom and
# the restricted model 43, so the difference test has 1.
result <- ss_buc_chisq_diff(chisq_observed = 9.5, N = 250, df_full = 42,
  df_restricted = 43, alpha_prior = .05, alpha_planned = .05,
  assurance = .80, desired_power = .80)
result

# A saturated full model means the comparison is an omnibus test of model
# fit, which this method does not describe, and the function says so:
try(ss_buc_chisq_diff(chisq_observed = 9.5, N = 250, df_full = 0,
  df_restricted = 43))

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .95):
try(ss_buc_chisq_diff(chisq_observed = 9.5, N = 250, df_full = 42,
  df_restricted = 43, assurance = .99))
```

ss_buc_correlation	<i>Necessary sample size to reach desired power for a Pearson correlation using a publication bias and uncertainty correction procedure</i>
--------------------	---

Description

ss_buc_correlation returns the necessary total sample size to achieve a desired level of statistical power for a test of a Pearson correlation in a planned study, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

Usage

```
ss_buc_correlation(
  r_observed,
  N,
  t_observed,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

r_observed	Observed Pearson correlation from a previous study used to plan sample size for a planned study. Either sign is accepted; only the magnitude enters the correction. Supply either r_observed or t_observed, not both.
N	Total sample size of the previous study.
t_observed	Observed t statistic for the correlation from the previous study, an alternative to r_observed.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in `ss_buc_correlation` uses the observed correlation and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then used to calculate the necessary total sample size to achieve the desired level of power in the planned study.

The observed correlation r with sample size N gives $t = r\sqrt{(N-2)/(1-r^2)}$ on $N-2$ degrees of freedom, and the planner works from that t . Supplying `t_observed` directly instead of `r_observed` is equivalent.

The approach uses a likelihood function of a truncated distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of the test statistic; the denominator is the power of the test, which serves to truncate the distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.) Which density belongs in that numerator is the one substantive difference between this planner and `ss_buc_reg_coef`, and it is described next.

Both variables are sampled, and that changes the distribution. The test of a correlation is arithmetically the test of the slope in a simple regression, so it is tempting to plan it as one. That is exact only when the predictor is fixed by design, as it is in an experiment where the researcher sets the values of X . A correlation study samples *both* variables. Conditional on the sampled X , the statistic is noncentral F , but its noncentrality parameter is itself random: it is proportional to

the sampled spread of X , which follows a chi-square distribution on $N - 1$ degrees of freedom. The marginal distribution is therefore a chi-square mixture of noncentral F distributions, which is more dispersed than any single noncentral F , and treating the statistic as noncentral F overstates the power of the planned study by up to about .04.

`ss_buc_correlation` uses the mixture, which has an exact closed form (the Poisson weights of the noncentral F become negative binomial weights), for both the correction and the planned study's power. The returned sample size is therefore exact for bivariate normal data rather than a few participants light, and it is typically 2 to 5 participants larger than the fixed-predictor calculation, more when the plan is small. Near the assurance ceiling the difference reverses and the mixture calls for a slightly smaller sample. If your predictor really is fixed by design, the analysis is a regression slope rather than a correlation, and `ss_buc_reg_coef` with $p = 1$ is the planner for it.

The largest assurance the prior result can support is the same in either frame, since at a zero effect the mixture is the central F distribution.

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, `assurance = .5` means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting `assurance = .5` (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, `assurance > .5` can be selected, which corresponds to selecting the noncentrality parameter associated with the $(1 - \text{assurance})$ quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior = .20` would reflect less severe publication bias than the default of .05. In essence, setting `alpha_prior` at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior = 1` and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

The observed correlation may be entered with either sign, since the publication rule the correction assumes is two-sided and only the magnitude enters the computation.

The returned `ncp_adjusted` is the noncentrality parameter the adjusted correlation implies at the prior study's sample size, the same scale the other planners report it on. The adjusted correlation itself is $\sqrt{\text{ncp_adjusted} / (\text{ncp_adjusted} + N)}$.

Value

An object of class `bucss_power`: a `data.frame` with a character term column and a numeric value column, in the same shape as the sibling package `DMAR`'s planners. The design results come first: the necessary total sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:[10.1037/met0000520](https://doi.org/10.1037/met0000520)
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:[10.1177/0956797617723724](https://doi.org/10.1177/0956797617723724)
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.
- Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

See Also

[ss_buc_reg_coef](#) for a coefficient in a model with more than one predictor.

Examples

```
result <- ss_buc_correlation(r_observed = .35, N = 100, alpha_prior = .05,
  alpha_planned = .05, assurance = .80, desired_power = .80)
result
```

```
# The equivalent call from the observed t statistic:
ss_buc_correlation(t_observed = .35 * sqrt(98 / (1 - .35^2)), N = 100)

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance. The correlation
# above supports .99, so a weaker prior is needed to reach the ceiling
# (here near .86):
try(ss_buc_correlation(r_observed = .27, N = 100, assurance = .95))
```

ss_buc_factorial_anova

Necessary sample size to reach desired power for a two-way between-subjects ANOVA using a publication bias and uncertainty correction procedure

Description

ss_buc_factorial_anova returns the necessary per-cell sample size to achieve a desired level of statistical power for a planned study testing an omnibus main effect or interaction in a two-way fully between-subjects ANOVA, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

For a single between-subjects factor, use [ss_buc_one_way_anova](#); for a design with more than two factors, or for a single-degree-of-freedom contrast, use [ss_buc_factorial_anova_general](#).

Usage

```
ss_buc_factorial_anova(
  F_observed,
  N,
  levels_A,
  levels_B,
  effect = c("factor_A", "factor_B", "interaction"),
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

F_observed	Observed <i>F</i> value from a previous study used to plan sample size for a planned study.
------------	---

N	Total sample size of the previous study.
levels_A	Number of levels for Factor A.
levels_B	Number of levels for Factor B.
effect	Effect of most interest to the planned study: main effect of A (factor_A), main effect of B (factor_B), or interaction (interaction). Case-insensitive shorthand is also accepted but not recommended in scripts: "A" (or "a") for factor_A, "B" for factor_B, and "AxB" (or "AB") for interaction.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in `ss_buc_factorial_anova` uses the observed F value and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then used to calculate the necessary per-cell sample size to achieve the desired level of power in the planned study.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. Thus, the ratio of the numerator and the denominator is a truncated noncentral F distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, assurance = .5 means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the (1 - assurance) quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior = .20` would reflect less severe publication bias than the default of .05. In essence, setting `alpha_prior` at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior = 1` and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

`ss_buc_factorial_anova` assumes that the planned study will have equal n . Unequal n in the previous study is handled in the following way for between-subjects ANOVA designs. If the user enters an N not equally divisible by the number of cells, the function calculates n by dividing N by the number of cells and both rounding up and rounding down the result, effectively assuming equal n . The suggested sample size for the planned study is calculated using both of these values of n , and the function returns the larger of these two suggestions, to be conservative. The adjusted noncentrality parameter that is output is the lower of the two possibilities, again, to be conservative. Although equal- n previous studies are preferable, this approach will work well as long as the cell sizes are only slightly discrepant.

Value

An object of class `bucss_power`: a `data.frame` with a character `term` column and a numeric `value` column, in the same shape as the sibling package `DMAR`'s planners. The design results come first: the necessary per-cell sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the effect tested, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.
- Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

Examples

```
result <- ss_buc_factorial_anova(F_observed = 5, N = 120, levels_A = 2,
  levels_B = 3, effect = "factor_B", alpha_prior = .05, alpha_planned = .05,
  assurance = .80, desired_power = .80)
result

# Pull the planned per-cell sample size, the implied total sample size, and
# the adjusted noncentrality parameter out of the result:
result$value[result$term == "necessary_n_per_cell"] # per cell
result$value[result$term == "total_N"]             # total across all cells
result$value[result$term == "ncp_adjusted"]

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .83):
try(ss_buc_factorial_anova(F_observed = 5, N = 120, levels_A = 2,
  levels_B = 3, effect = "factor_B", assurance = .95))
```

ss_buc_factorial_anova_general

Necessary sample size to reach desired power for a between-subjects ANOVA with any number of factors using a publication bias and uncertainty correction procedure

Description

`ss_buc_factorial_anova_general` returns the necessary per-cell sample size to achieve a desired level of statistical power for a planned study testing any type of effect (omnibus or contrast) using a fully between-subjects ANOVA with any number of factors, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

For the common one-way and two-way omnibus designs, `ss_buc_one_way_anova` and `ss_buc_factorial_anova` take the factor levels directly; this general form instead takes the degrees of freedom, so it covers more factors and single-degree-of-freedom contrasts.

Usage

```
ss_buc_factorial_anova_general(
  F_observed,
  N,
  cells,
  df_numerator,
  df_denominator,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

<code>F_observed</code>	Observed F value from a previous study used to plan sample size for a planned study.
<code>N</code>	Total sample size of the previous study.
<code>cells</code>	Number of cells for the design (the product of the number of levels of each factor).
<code>df_numerator</code>	Numerator degrees of freedom for the effect of interest.
<code>df_denominator</code>	Denominator (error) degrees of freedom for the effect of interest in the prior study. This equals $N - \text{cells}$ when the cell means are the only estimated parameters; supply a smaller value to account for covariates or other nuisance parameters. It cannot exceed $N - \text{cells}$.
<code>alpha_prior</code>	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
<code>alpha_planned</code>	Alpha level α assumed for the planned study.
<code>assurance</code>	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level ($\text{assurance} > .5$ corrects for uncertainty; $\text{assurance} < .5$ is not recommended). Enter it as a proportion in (0,

1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.

desired_power Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in `ss_buc_factorial_anova_general` uses the observed F value and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then used to calculate the necessary per-cell sample size to achieve the desired level of power in the planned study.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. Thus, the ratio of the numerator and the denominator is a truncated noncentral F distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

The denominator degrees of freedom you supply describe the prior study's F test for the effect of interest. In a pure cell-means model the residual degrees of freedom equal N minus the number of cells; supplying a smaller value accommodates additional estimated parameters, such as covariates or blocking factors. The difference, $(N - \text{cells}) - \text{df_denominator}$, is treated as the number of such nuisance parameters and is carried through to the planned study, whose denominator degrees of freedom are reduced by the same amount. Supplying $\text{df_denominator} = N - \text{cells}$ reproduces the case with no such extra parameters, and a smaller df_denominator calls for a larger sample size. df_denominator cannot exceed $N - \text{cells}$.

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, $\text{assurance} = .5$ means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting $\text{assurance} = .5$ (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, $\text{assurance} > .5$ can be selected, which corresponds to selecting the noncentrality parameter associated with the $(1 - \text{assurance})$ quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), α_{prior} can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting $\alpha_{\text{prior}} = .20$ would reflect less severe publication bias than the default of .05. In essence, setting α_{prior} at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior` = 1 and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

`ss_buc_factorial_anova_general` assumes that the planned study will have equal n . Unequal n in the previous study is handled in the following way for between-subjects ANOVA designs. If the user enters an N not equally divisible by the number of cells, the function calculates n by dividing N by the number of cells and both rounding up and rounding down the result, effectively assuming equal n . The suggested sample size for the planned study is calculated using both of these values of n , and the function returns the larger of these two suggestions, to be conservative. The adjusted noncentrality parameter that is output is the lower of the two possibilities, again, to be conservative. Although equal- n previous studies are preferable, this approach will work well as long as the cell sizes are only slightly discrepant.

Value

An object of class `bucss_power`: a `data.frame` with a character `term` column and a numeric value column, in the same shape as the sibling package `DMAR`'s planners. The design results come first: the necessary per-cell sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see `Details`); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.
- Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

Examples

```
# With df_denominator = N - cells the general form reproduces the two-way
# omnibus result from ss_buc_factorial_anova (here 120 - 6 = 114).
result <- ss_buc_factorial_anova_general(F_observed = 5, N = 120, cells = 6,
  df_numerator = 2, df_denominator = 114, alpha_prior = .05,
  alpha_planned = .05, assurance = .80, desired_power = .80)
result

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .83):
try(ss_buc_factorial_anova_general(F_observed = 5, N = 120, cells = 6,
  df_numerator = 2, df_denominator = 114, assurance = .95))
```

ss_buc_independent_t	<i>Necessary sample size to reach desired power for an independent t test (or standardized mean difference) using a publication bias and uncertainty correction procedure</i>
----------------------	---

Description

ss_buc_independent_t returns the necessary per-group sample size to achieve a desired level of statistical power for a planned study using an independent *t* test, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

ss_buc_smd is an alias for ss_buc_independent_t: the two are the same function. The two-group *t* test and the standardized mean difference (Cohen's *d*) describe the same comparison, so the same planner serves both framings. Prefer the test-specific name (ss_buc_independent_t) when the

prior t was computed from raw (unstandardized) data, since the planner works directly from the observed t .

Usage

```
ss_buc_independent_t(
  t_observed,
  n,
  N,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)

ss_buc_smd(
  t_observed,
  n,
  N,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

t_observed	Observed t value from a previous study used to plan sample size for a planned study. Either sign is accepted: the publication rule is two-sided, so only the magnitude enters the correction (see Details).
n	Per group sample size (if equal) or the two group sample sizes of the previous study (enter either a single number or a vector of length 2).
N	Total sample size of the previous study, assumed equal across groups if specified.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in `ss_buc_independent_t` uses the observed t value and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then used to calculate the necessary per-group sample size to achieve the desired level of power in the planned study.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. In the two-group case, this formula reduces to the density of a truncated noncentral t distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, assurance = .5 means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the $(1 - \text{assurance})$ quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior` = .20 would reflect less severe publication bias than the default of .05. In essence, setting `alpha_prior` at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior` = 1 and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted

noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

The observed t may be entered with either sign. The publication rule the correction assumes is two-sided and the truncated likelihood is symmetric in the sign of t , so only the magnitude enters the computation: `t_observed = -3` plans exactly the sample size that `t_observed = 3` plans. The sign of a two-group t records which group was subtracted from which, an arbitrary coding choice rather than evidence. When planning a replication, confirm that the magnitude belongs to an effect in the direction the planned study is designed to detect (Anderson & Kelley, 2024): the correction adjusts the size of the prior effect, not its direction.

If you are working from a standardized mean difference (Cohen's d) rather than a t statistic, convert it before calling: for two independent groups $t = d / \sqrt{1/n_1 + 1/n_2}$, which equals $d\sqrt{n/2}$ when the per-group sample sizes are equal.

`ss_buc_independent_t` assumes that the planned study will have equal n . Unequal n in the previous study is handled in the following way for the independent t . If the user enters an odd value for N , no information is available on the exact group sizes. The function calculates n by dividing N by 2 and both rounding up and rounding down the result, thus assuming equal n . The suggested sample size for the planned study is calculated using both of these values of n , and the function returns the larger of these two suggestions, to be conservative. If the user enters a vector for n with two different values, specific information is available on the exact group sizes. n is calculated as the harmonic mean of these two values (a measure of effective sample size). Again, this is rounded both up and down. The suggested sample size for the planned study is calculated using both of these values of n , and the function returns the larger of these two suggestions, to be conservative. The adjusted noncentrality parameter that is output is the lower of the two possibilities, again, to be conservative. When the individual group sizes of an unequal- n previous study are known, we highly encourage entering these explicitly, especially if the sample sizes are quite discrepant, as this allows for the greatest precision in estimating an appropriate planned study n .

Value

An object of class `bucss_power`: a `data.frame` with a character `term` column and a numeric `value` column, in the same shape as the sibling package `DMAR`'s planners. The design results come first: the necessary per-group sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724

Anderson, S. F., & Maxwell, S. E. (2017). Addressing the ‘replication crisis’: Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.

Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

Examples

```
result <- ss_buc_independent_t(t_observed = 3, n = 20, alpha_prior = .05,
  alpha_planned = .05, assurance = .80, desired_power = .80)
result
result$value[result$term == "necessary_n_per_group"]

# ss_buc_smd is the same function under an effect size name
ss_buc_smd(t_observed = 3, n = 20)

# Asking for more assurance than the prior result can support stops with an
# informative error that reports the largest workable assurance (here near
# .90, the value 1 - p/alpha_prior for this prior result):
try(ss_buc_independent_t(t_observed = 3, n = 20, assurance = .95))
```

ss_buc_manova	Necessary sample size to reach desired power for a two-group multivariate comparison using a publication bias and uncertainty correction procedure
---------------	--

Description

ss_buc_manova returns the necessary per-group sample size to achieve a desired level of statistical power for a planned study comparing two groups on several outcome variables at once (Hotelling’s T^2 , equivalently a one-way multivariate analysis of variance with two groups), based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

Usage

```
ss_buc_manova(
  F_observed,
  N,
  p_variables,
  T2_observed,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

F_observed	Observed F value for the multivariate comparison from a previous study. Supply either F_observed or T2_observed, not both.
N	Total sample size of the previous study (both groups combined).
p_variables	Number of outcome variables compared.
T2_observed	Observed Hotelling's T^2 from the previous study, an alternative to F_observed.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

Scope. The correction this package implements works on a single test statistic with an exact non-central F distribution. A multivariate hypothesis has one only when its rank is one, which is the case for a comparison of two groups: Hotelling's T^2 on p outcome variables with total sample size N maps exactly to $F = T^2(N - p - 1)/(p(N - 2))$ on p and $N - p - 1$ degrees of freedom. Supply either the F or the T^2 . For a hypothesis of rank greater than one (three or more groups, say) the multivariate noncentrality is a set of eigenvalues rather than a single number, the usual test statistics

have only approximate F forms, and this correction does not apply; `ss_buc_manova` will not accept such a design.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. Thus, the ratio of the numerator and the denominator is a truncated noncentral F distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, `assurance = .5` means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting `assurance = .5` (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, `assurance > .5` can be selected, which corresponds to selecting the noncentrality parameter associated with the $(1 - \text{assurance})$ quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior = .20` would reflect less severe publication bias than the default of .05. In essence, setting `alpha_prior` at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior = 1` and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

`ss_buc_manova` assumes that the planned study will have equal n per group. If the prior N is odd, the per-group size is rounded both up and down, the planned sample size is computed under each, and the larger of the two is returned, with the smaller of the two adjusted noncentrality parameters, to be conservative.

Value

An object of class `bucss_power`: a `data.frame` with a character term column and a numeric value column, in the same shape as the sibling package `DMAR`'s planners. The design results come first: the necessary per-group sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study non-centrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.
- Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

Examples

```
result <- ss_buc_manova(F_observed = 4.5, N = 80, p_variables = 3,
  alpha_prior = .05, alpha_planned = .05, assurance = .80,
  desired_power = .80)
result

# The equivalent call from Hotelling's T squared:
ss_buc_manova(T2_observed = 4.5 * 3 * 78 / 76, N = 80, p_variables = 3)

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .88):
try(ss_buc_manova(F_observed = 4.5, N = 80, p_variables = 3,
  assurance = .95))
```

ss_buc_mixed_anova	<i>Necessary sample size to reach desired power for a two-factor split-plot (mixed) ANOVA using a publication bias and uncertainty correction procedure</i>
--------------------	---

Description

ss_buc_mixed_anova returns the necessary per between-subjects cell sample size to achieve a desired level of statistical power for a planned study testing an omnibus effect using a two-factor split-plot (mixed) ANOVA, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

Usage

```
ss_buc_mixed_anova(
  F_observed,
  N,
  levels_between,
  levels_within,
  effect = c("between", "within", "interaction"),
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

F_observed	Observed F value from a previous study used to plan sample size for a planned study.
N	Total sample size of the previous study.
levels_between	Number of levels for the between-subjects factor.
levels_within	Number of levels for the within-subjects factor.
effect	Effect of most interest to the planned study: between-subjects main effect (between), within-subjects main effect (within), or interaction (interaction). Case-insensitive shorthand is also accepted but not recommended in scripts: "bs" for between, "ws" for within, and "BxW" for interaction.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.

assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in `ss_buc_mixed_anova` uses the observed F value and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then used to calculate the necessary per between-subjects cell sample size to achieve the desired level of power in the planned study.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. Thus, the ratio of the numerator and the denominator is a truncated noncentral F distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, assurance = .5 means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the $(1 - \text{assurance})$ quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior` = .20 would reflect less severe publication bias than the default of .05. In essence, setting `alpha_prior` at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`;

at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior` = 1 and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

`ss_buc_mixed_anova` assumes that the planned study will have equal n . Unequal n in the previous study is handled in the following way for split-plot designs. If the user enters an N not equally divisible by the number of between-subjects cells, the function calculates n by dividing N by the number of cells and both rounding up and rounding down the result, effectively assuming equal n . The suggested sample size for the planned study is calculated using both of these values of n , and the function returns the larger of these two suggestions, to be conservative. The adjusted noncentrality parameter that is output is the lower of the two possibilities, again, to be conservative. Although equal- n previous studies are preferable, this approach will work well as long as the cell sizes are only slightly discrepant.

`ss_buc_mixed_anova` assumes sphericity for the within-subjects effects.

Value

An object of class `bucss_power`: a `data.frame` with a character `term` column and a numeric value column, in the same shape as the sibling package `DMAR's planners`. The design results come first: the necessary per between-subjects cell sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the effect tested, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724

Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.

Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

Examples

```
result <- ss_buc_mixed_anova(F_observed = 5, N = 60, levels_between = 2,
  levels_within = 3, effect = "within", alpha_prior = .05,
  alpha_planned = .05, assurance = .80, desired_power = .80)
result

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .83):
try(ss_buc_mixed_anova(F_observed = 5, N = 60, levels_between = 2,
  levels_within = 3, effect = "within", assurance = .95))
```

ss_buc_mixed_anova_general

Necessary sample size to reach desired power for a split-plot (mixed) ANOVA with any number of factors using a publication bias and uncertainty correction procedure

Description

ss_buc_mixed_anova_general returns the necessary per-group sample size to achieve a desired level of statistical power for a planned study testing any type of effect (omnibus, contrast) using a split-plot (mixed) ANOVA with any number of factors, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

Usage

```
ss_buc_mixed_anova_general(
  F_observed,
  N,
  df_numerator,
```

```

num_groups,
effect = c("between_only", "within_only", "between_within"),
df_num_within,
alpha_prior = 0.05,
alpha_planned = 0.05,
assurance = 0.8,
desired_power = 0.8
)

```

Arguments

F_observed	Observed F value from a previous study used to plan sample size for a planned study.
N	Total sample size of the previous study.
df_numerator	Numerator degrees of freedom for the effect of interest.
num_groups	Number of distinct groups (product of the number of levels of the between-subjects factors).
effect	Effect of interest: involves only between-subjects effects (between_only), involves only within-subjects effects (within_only), or involves both between and within effects (between_within).
df_num_within	Numerator degrees of freedom for only the within-subjects components of the effect of interest. Only needed when effect = "between_within".
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in `ss_buc_mixed_anova_general` uses the observed F value and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then

used to calculate the necessary per-group sample size to achieve the desired level of power in the planned study.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. Thus, the ratio of the numerator and the denominator is a truncated noncentral F distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, assurance = .5 means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the $(1 - \text{assurance})$ quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior` = .20 would reflect less severe publication bias than the default of .05. In essence, setting `alpha_prior` at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior` = 1 and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

`ss_buc_mixed_anova_general` assumes that the planned study will have equal n . Unequal n in the previous study is handled in the following way for split-plot designs. If the user enters an N not equally divisible by the number of between-subjects cells, the function calculates n by dividing N by the number of cells and both rounding up and rounding down the result, effectively assuming equal n . The suggested sample size for the planned study is calculated using both of these values of n , and the function returns the larger of these two suggestions, to be conservative. The adjusted

noncentrality parameter that is output is the lower of the two possibilities, again, to be conservative. Although equal-n previous studies are preferable, this approach will work well as long as the cell sizes are only slightly discrepant.

ss_buc_mixed_anova_general assumes sphericity for the within-subjects effects.

Value

An object of class `bucss_power`: a `data.frame` with a character term column and a numeric value column, in the same shape as the sibling package `DMAR`'s planners. The design results come first: the necessary per-group sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the effect tested, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.
- Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

Examples

```
result <- ss_buc_mixed_anova_general(F_observed = 5, N = 90, df_numerator = 2,
  num_groups = 3, effect = "between_only",
  alpha_prior = .05, alpha_planned = .05, assurance = .80, desired_power = .80)
result

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .82):
try(ss_buc_mixed_anova_general(F_observed = 5, N = 90, df_numerator = 2,
```

```
num_groups = 3, effect = "between_only",
assurance = .95))
```

ss_buc_one_sample_t	<i>Necessary sample size to reach desired power for a one-sample t test using a publication bias and uncertainty correction procedure</i>
---------------------	--

Description

ss_buc_one_sample_t returns the necessary total sample size to achieve a desired level of statistical power for a planned study using a one-sample t test, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

Usage

```
ss_buc_one_sample_t(
  t_observed,
  N,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

t_observed	Observed t value from a previous study used to plan sample size for a planned study. Either sign is accepted: the publication rule is two-sided, so only the magnitude enters the correction.
N	Total sample size of the previous study.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.

`desired_power` Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

The one-sample t test and the dependent (paired) t test are the same test applied to a single column of scores: the statistic has $N - 1$ degrees of freedom and a noncentrality parameter of $d\sqrt{N}$. `ss_buc_one_sample_t` therefore performs exactly the computation `ss_buc_paired_t` performs, and reports it with a one-sample design label and a total sample size rather than a number of pairs. See `ss_buc_paired_t` for the full description of the method, the meaning of assurance, the role of `alpha_prior`, and what happens when the corrected noncentrality parameter is zero.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

The observed t may be entered with either sign; the publication rule the correction assumes is two-sided, so only the magnitude enters the computation.

If you are working from a standardized mean (Cohen's d) rather than a t statistic, convert it before calling: $t = d\sqrt{N}$.

Value

An object of class `bucss_power`: a `data.frame` with a character term column and a numeric value column, in the same shape as the sibling package DMAR's planners. The design results come first: the necessary total sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724

Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.

Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

See Also

[ss_buc_paired_t](#), which performs the identical computation for a paired design.

Examples

```
result <- ss_buc_one_sample_t(t_observed = 3, N = 40, alpha_prior = .05,
  alpha_planned = .05, assurance = .80, desired_power = .80)
result

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .90):
try(ss_buc_one_sample_t(t_observed = 3, N = 40, assurance = .95))
```

ss_buc_one_way_anova	<i>Necessary sample size to reach desired power for a one-way between-subjects ANOVA using a publication bias and uncertainty correction procedure</i>
----------------------	--

Description

ss_buc_one_way_anova returns the necessary per-group sample size to achieve a desired level of statistical power for a planned study testing the omnibus effect of a single between-subjects factor, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

For a fully between-subjects design with two crossed factors, use [ss_buc_factorial_anova](#); for any other between-subjects design (more factors, or a single-degree-of-freedom contrast), use [ss_buc_factorial_anova_ge](#).

Usage

```
ss_buc_one_way_anova(
  F_observed,
  N,
  levels_A,
```

```

alpha_prior = 0.05,
alpha_planned = 0.05,
assurance = 0.8,
desired_power = 0.8
)

```

Arguments

F_observed	Observed F value from a previous study used to plan sample size for a planned study.
N	Total sample size of the previous study.
levels_A	Number of levels (groups) for the single between-subjects factor.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level ($\text{assurance} > .5$ corrects for uncertainty; $\text{assurance} < .5$ is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in `ss_buc_one_way_anova` uses the observed F value and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then used to calculate the necessary per-group sample size to achieve the desired level of power in the planned study.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. Thus, the ratio of the numerator and the denominator is a truncated noncentral F distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, $\text{assurance} = .5$ means that power will be above the desired level half of the time, but below the desired level the other half of the time.

Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the (1 - assurance) quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), alpha_prior can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting alpha_prior = .20 would reflect less severe publication bias than the default of .05. In essence, setting alpha_prior at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is alpha_prior; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting alpha_prior at a value greater than .05, which raises the ceiling. Note that alpha_prior is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and alpha_prior is the only remaining lever. It is possible to correct for uncertainty only by setting alpha_prior = 1 and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned actual_power is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, actual_power is never below the desired_power that was requested and is usually a little above it.

ss_buc_one_way_anova assumes that the planned study will have equal n . Unequal n in the previous study is handled in the following way for between-subjects ANOVA designs. If the user enters an N not equally divisible by the number of groups, the function calculates n by dividing N by the number of groups and both rounding up and rounding down the result, effectively assuming equal n . The suggested sample size for the planned study is calculated using both of these values of n , and the function returns the larger of these two suggestions, to be conservative. The adjusted noncentrality parameter that is output is the lower of the two possibilities, again, to be conservative. Although equal- n previous studies are preferable, this approach will work well as long as the group sizes are only slightly discrepant.

Value

An object of class bucss_power: a data.frame with a character term column and a numeric value column, in the same shape as the sibling package DMAR's planners. The design results come first: the necessary per-group sample size for the planned study, named for its unit (necessary_n_per_group, necessary_n_per_cell, or necessary_N); for the per-group and per-cell designs, total_N, the implied total sample size; actual_power, the conservative achieved power at the returned sample

size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study non-centrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.
- Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

Examples

```
result <- ss_buc_one_way_anova(F_observed = 5, N = 120, levels_A = 4,
  alpha_prior = .05, alpha_planned = .05, assurance = .80, desired_power = .80)
result

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .94):
try(ss_buc_one_way_anova(F_observed = 5, N = 120, levels_A = 4,
  assurance = .95))
```

ss_buc_paired_t	<i>Necessary sample size to reach desired power for a dependent (paired) t test (or standardized mean difference) using a publication bias and uncertainty correction procedure</i>
-----------------	---

Description

ss_buc_paired_t returns the necessary sample size (the number of pairs) to achieve a desired level of statistical power for a planned study using a dependent t test, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

ss_buc_smd_paired is an alias for ss_buc_paired_t: the two are the same function. The paired t test and the standardized mean difference of the difference scores describe the same comparison, so the same planner serves both framings. Prefer the test-specific name (ss_buc_paired_t) when the prior t was computed from raw (unstandardized) data, since the planner works directly from the observed t .

Usage

```
ss_buc_paired_t(
  t_observed,
  N,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)

ss_buc_smd_paired(
  t_observed,
  N,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

t_observed	Observed t value from a previous study used to plan sample size for a planned study. Either sign is accepted: the publication rule is two-sided, so only the magnitude enters the correction (see Details).
N	Total sample size (the number of pairs) of the previous study.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0,

1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.

desired_power Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in `ss_buc_paired_t` uses the observed t value and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then used to calculate the necessary sample size to achieve the desired level of power in the planned study.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. In the two-group case, this formula reduces to the density of a truncated noncentral t distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, assurance = .5 means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the $(1 - \text{assurance})$ quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior` = .20 would reflect less severe publication bias than the default of .05. In essence, setting `alpha_prior` at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the

prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior = 1` and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

The observed t may be entered with either sign. The publication rule the correction assumes is two-sided and the truncated likelihood is symmetric in the sign of t , so only the magnitude enters the computation: `t_observed = -3` plans exactly the sample size that `t_observed = 3` plans. The sign of a paired t records the direction the within-pair difference was taken, an arbitrary coding choice rather than evidence. When planning a replication, confirm that the magnitude belongs to an effect in the direction the planned study is designed to detect (Anderson & Kelley, 2024): the correction adjusts the size of the prior effect, not its direction.

If you are working from a standardized mean difference (Cohen's d_z) of the difference scores rather than a t statistic, convert it before calling: for a paired design $t = d_z \sqrt{N}$, where N is the number of pairs.

Value

An object of class `bucss_power`: a data.frame with a character `term` column and a numeric `value` column, in the same shape as the sibling package DMAR's planners. The design results come first: the necessary number of pairs for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.

Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

Examples

```
result <- ss_buc_paired_t(t_observed = 3, N = 40, alpha_prior = .05,
  alpha_planned = .05, assurance = .80, desired_power = .80)
result

# ss_buc_smd_paired is the same function under an effect size name
ss_buc_smd_paired(t_observed = 3, N = 40)

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .90):
try(ss_buc_paired_t(t_observed = 3, N = 40, assurance = .95))
```

ss_buc_R2

Necessary sample size to reach desired power for a test of model R^2 in a multiple regression using a publication bias and uncertainty correction procedure

Description

ss_buc_R2 returns the necessary total sample size to achieve a desired level of statistical power for a test of model R^2 in a planned study using multiple regression, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

Usage

```
ss_buc_R2(
  F_observed,
  N,
  p,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

F_observed	Observed F value from a previous study used to plan sample size for a planned study.
N	Total sample size of the previous study.
p	Number of predictors; be sure to include any product terms or polynomials that are in the model.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in ss_buc_R2 uses the observed F value and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then used to calculate the necessary total sample size to achieve the desired level of power in the planned study.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. Thus, the ratio of the numerator and the denominator is a truncated noncentral F distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, assurance = .5 means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the (1 - assurance) quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior = .20` would reflect less severe publication bias than the default of `.05`. In essence, setting `alpha_prior` at `.20` assumes that studies with p values less than `.20` are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., `.8` instead of `.95`), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than `.05`, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of `.5`, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior = 1` and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

Value

An object of class `bucss_power`: a `data.frame` with a character term column and a numeric value column, in the same shape as the sibling package `DMAR`'s planners. The design results come first: the necessary total sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724

Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.

Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

Anderson, S. F. (2021). Using prior information to plan appropriately powered regression studies: A tutorial using BUCSS. *Psychological Methods*, 26, 513–526. doi:10.1037/met0000366

Examples

```
result <- ss_buc_R2(F_observed = 5, N = 150, p = 4, alpha_prior = .05,
  alpha_planned = .05, assurance = .80, desired_power = .80)
result

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .98):
try(ss_buc_R2(F_observed = 5, N = 150, p = 4, assurance = .99))
```

ss_buc_reg_coef	<i>Necessary sample size to reach desired power for a single coefficient in a multiple regression using a publication bias and uncertainty correction procedure</i>
-----------------	---

Description

ss_buc_reg_coef returns the necessary total sample size to achieve a desired level of statistical power for a single regression coefficient in a planned study using multiple regression, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

Usage

```
ss_buc_reg_coef(
  t_observed,
  N,
  p,
  alpha_prior = 0.05,
```

```

alpha_planned = 0.05,
assurance = 0.8,
desired_power = 0.8
)

```

Arguments

t_observed	Observed t value from a previous study used to plan sample size for a planned study. Either sign is accepted: the publication rule is two-sided, so only the magnitude enters the correction (see Details).
N	Total sample size of the previous study.
p	Number of predictors; be sure to include any product terms or polynomials that are in the model.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in `ss_buc_reg_coef` uses the observed t value and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then used to calculate the necessary total sample size to achieve the desired level of power in the planned study.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. In the single predictor case, this formula reduces to the density of a truncated noncentral t distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, assurance = .5 means that power will be above the desired level half of the time, but below the desired level the other half of the time.

Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the (1 - assurance) quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), alpha_prior can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting alpha_prior = .20 would reflect less severe publication bias than the default of .05. In essence, setting alpha_prior at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is alpha_prior; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting alpha_prior at a value greater than .05, which raises the ceiling. Note that alpha_prior is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and alpha_prior is the only remaining lever. It is possible to correct for uncertainty only by setting alpha_prior = 1 and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned actual_power is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, actual_power is never below the desired_power that was requested and is usually a little above it.

The coefficient's observed t may be entered with either sign. The publication rule the correction assumes is two-sided and the statistic enters on the F scale as t^2 , so only the magnitude affects the computation: $t_{\text{observed}} = -3$ plans exactly the sample size that $t_{\text{observed}} = 3$ plans. When planning a replication, confirm that the magnitude belongs to a coefficient in the direction the planned study is designed to detect (Anderson & Kelley, 2024): the correction adjusts the size of the prior effect, not its direction.

Value

An object of class bucss_power: a data.frame with a character term column and a numeric value column, in the same shape as the sibling package DMAR's planners. The design results come first: the necessary total sample size for the planned study, named for its unit (necessary_n_per_group, necessary_n_per_cell, or necessary_N); for the per-group and per-cell designs, total_N, the implied total sample size; actual_power, the conservative achieved power at the returned sample size (see Details); and ncp_adjusted, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support

(the assurance ceiling), and the planned test's degrees of freedom (df_effect and df_error) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.
- Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.
- Anderson, S. F. (2021). Using prior information to plan appropriately powered regression studies: A tutorial using BUCSS. *Psychological Methods*, 26, 513–526. doi:10.1037/met0000366

Examples

```
result <- ss_buc_reg_coef(t_observed = 3, N = 150, p = 3, alpha_prior = .05,
  alpha_planned = .05, assurance = .80, desired_power = .80)
result

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .93):
try(ss_buc_reg_coef(t_observed = 3, N = 150, p = 3, assurance = .95))
```

ss_buc_reg_joint

Necessary sample size to reach desired power for a joint test of multiple predictors in a multiple regression using a publication bias and uncertainty correction procedure

Description

ss_buc_reg_joint returns the necessary total sample size to achieve a desired level of statistical power for a joint test of multiple predictors in a planned study using multiple regression, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

Usage

```
ss_buc_reg_joint(
  F_observed,
  N,
  p,
  p_joint,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

F_observed	Observed F value from a previous study used to plan sample size for a planned study.
N	Total sample size of the previous study.
p	Number of predictors; be sure to include any product terms or polynomials that are in the model.
p_joint	Number of predictors tested jointly for significance.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in `ss_buc_reg_joint` uses the observed F value and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then used to calculate the necessary total sample size to achieve the desired level of power in the planned study.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. Thus, the ratio of the numerator and the denominator is a truncated noncentral F distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, assurance = .5 means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the $(1 - \text{assurance})$ quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior` = .20 would reflect less severe publication bias than the default of .05. In essence, setting `alpha_prior` at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior` = 1 and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted

noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

Value

An object of class `bucss_power`: a `data.frame` with a character term column and a numeric value column, in the same shape as the sibling package `DMAR`'s planners. The design results come first: the necessary total sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.
- Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.
- Anderson, S. F. (2021). Using prior information to plan appropriately powered regression studies: A tutorial using BUCSS. *Psychological Methods*, 26, 513–526. doi:10.1037/met0000366

Examples

```
result <- ss_buc_reg_joint(F_observed = 5, N = 150, p = 4, p_joint = 2,
  alpha_prior = .05, alpha_planned = .05, assurance = .80, desired_power = .80)
result

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .84):
try(ss_buc_reg_joint(F_observed = 5, N = 150, p = 4, p_joint = 2,
  assurance = .95))
```

ss_buc_rm_anova	<i>Necessary sample size to reach desired power for a one or two-way within-subjects ANOVA using a publication bias and uncertainty correction procedure</i>
-----------------	--

Description

ss_buc_rm_anova returns the necessary total sample size to achieve a desired level of statistical power for a planned study testing an omnibus effect using a one or two-way fully within-subjects ANOVA, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

Usage

```
ss_buc_rm_anova(
  F_observed,
  N,
  levels_A,
  levels_B = NULL,
  effect = c("factor_A", "factor_B", "interaction"),
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

F_observed	Observed F value from a previous study used to plan sample size for a planned study.
N	Total sample size of the previous study.
levels_A	Number of levels for Factor A.
levels_B	Number of levels for Factor B, which is NULL if a single factor design.
effect	Effect of most interest to the planned study: main effect of A (factor_A), main effect of B (factor_B), or interaction (interaction). Case-insensitive shorthand is also accepted but not recommended in scripts: "A" (or "a") for factor_A, "B" for factor_B, and "AxB" (or "AB") for interaction.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.

assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in `ss_buc_rm_anova` uses the observed F value and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then used to calculate the necessary total sample size to achieve the desired level of power in the planned study.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. Thus, the ratio of the numerator and the denominator is a truncated noncentral F distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, assurance = .5 means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the $(1 - \text{assurance})$ quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior` = .20 would reflect less severe publication bias than the default of .05. In essence, setting `alpha_prior` at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are

available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior` = 1 and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

`ss_buc_rm_anova` assumes sphericity for the within-subjects effects.

Value

An object of class `bucss_power`: a `data.frame` with a character term column and a numeric value column, in the same shape as the sibling package `DMAR`'s planners. The design results come first: the necessary total sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the effect tested, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.
- Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

Examples

```
result <- ss_buc_rm_anova(F_observed = 5, N = 60, levels_A = 2, levels_B = 3,
  effect = "factor_B", alpha_prior = .05, alpha_planned = .05,
  assurance = .80, desired_power = .80)
result

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .83):
try(ss_buc_rm_anova(F_observed = 5, N = 60, levels_A = 2, levels_B = 3,
  effect = "factor_B", assurance = .95))
```

ss_buc_rm_anova_general

Necessary sample size to reach desired power for a within-subjects ANOVA with any number of factors using a publication bias and uncertainty correction procedure

Description

ss_buc_rm_anova_general returns the necessary total sample size to achieve a desired level of statistical power for a planned study testing any type of effect (omnibus, contrast) using a fully within-subjects ANOVA with any number of factors, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

Usage

```
ss_buc_rm_anova_general(
  F_observed,
  N,
  df_numerator,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

F_observed	Observed F value from a previous study used to plan sample size for a planned study.
N	Total sample size of the previous study.
df_numerator	Numerator degrees of freedom for the effect of interest.

alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty.

The approach implemented in *ss_buc_rm_anova_general* uses the observed F value and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is then used to calculate the necessary total sample size to achieve the desired level of power in the planned study.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. Thus, the ratio of the numerator and the denominator is a truncated noncentral F distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, assurance = .5 means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the (1 - assurance) quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), alpha_prior can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting alpha_prior = .20 would reflect less severe publication bias than the default of .05. In essence, setting alpha_prior at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect

cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior` = 1 and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

`ss_buc_rm_anova_general` assumes sphericity for the within-subjects effects.

Value

An object of class `bucss_power`: a `data.frame` with a character term column and a numeric value column, in the same shape as the sibling package DMAR's planners. The design results come first: the necessary total sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>) and Samantha F. Anderson (<samantha.f.anderson@asu.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.

Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

Examples

```
result <- ss_buc_rm_anova_general(F_observed = 6.5, N = 80, df_numerator = 1,
  alpha_prior = .05, alpha_planned = .05, assurance = .50, desired_power = .80)
result

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .74):
try(ss_buc_rm_anova_general(F_observed = 6.5, N = 80, df_numerator = 1,
  assurance = .95))
```

ss_buc_sensitivity	<i>Simulate the literature a plan came from and audit how often it works</i>
--------------------	--

Description

ss_buc_sensitivity takes a plan returned by any ss_buc_* function and asks the question the method’s assurance claim is really about: if the true effect were a particular value, and a literature published only the prior studies that reached significance, how often would a plan built this way actually reach the target power?

Usage

```
ss_buc_sensitivity(object, true_ncp, replications = 1000L, seed = NULL)
```

Arguments

object	A plan returned by any ss_buc_* function.
true_ncp	The assumed true noncentrality parameter of the prior study’s design, on the same scale as the ncp_adjusted row of object. Required; there is deliberately no default.
replications	Number of published prior studies to simulate. The default of 1000 gives a Monte Carlo standard error near .013 on the attainment rate; 10000 gives about .004.
seed	Optional random number seed. NULL (the default) leaves the random number stream alone. When supplied, the caller’s random number state is restored on exit.

Details

The bias and uncertainty correction rests on a claim about the long run. At assurance = .80, eight replications in ten should reach the target power. That claim is a statement about a whole literature, and a user planning one study never sees it exercised.

This function exercises it on the user's own numbers. It reads the design and the planning inputs off the plan you pass it, simulates prior studies of that design from a true effect *you* name, discards the ones a literature would not have published (those failing to reach significance at `alpha_prior`), runs the same planner on each survivor, and reports what happened.

You must name the true effect. There is no default, deliberately. Defaulting to the corrected estimate would be convenient and circular: it would ask the method to grade its own answer. `true_ncp` is on the same scale as the `ncp_adjusted` the plan reports, so the plan's own value is the natural anchor. Run the function at that value, at a fraction of it, and at a multiple of it; the interesting reading is how fast the attainment rate falls when the truth is smaller than the corrected estimate.

Reading the output. A plan attains the target when its sample size is at least the sample size the true effect actually requires, which is reported as `size_at_true_effect`. Two attainment rates are given, and the difference between them is not a rounding detail:

- `attainment` counts only the prior studies for which the method issued a plan. This is what a user experiences.
- `attainment_with_refusals` counts a refusal (the zero noncentrality parameter stop) as attaining, because declining to plan is the limiting case of an arbitrarily conservative plan. This is the quantity the assurance guarantee is about, and it should land on assurance.

When the refusal rate is high the first number sits well below assurance, and that is not a defect: it is the visible price of the method refusing rather than guessing. Its expected value is $(\text{assurance} - \text{refusal_rate}) / (1 - \text{refusal_rate})$.

What this can and cannot detect. The prior studies are drawn from the sampling distribution the method assumes for that design, so this is a check that the correction and the planned sample size are internally right, and a demonstration of what publication bias does. It cannot detect a violated distributional assumption, since the simulation shares that assumption. For `ss_buc_welch_t`, in particular, it holds the variance ratio fixed at the value you supplied.

Value

An object of class `bucss_sensitivity`, a `data.frame` with columns `term` and `value` carrying the assumed true parameter and the plan's own `ncp_adjusted` for comparison, the number of replications, the publication and refusal rates, the two attainment rates and the Monte Carlo standard error of the first, the sample size the true effect requires, and the quartiles of the sample sizes the simulated plans called for. The design, the sample size unit, the requested assurance, and the requested `desired_power` travel as attributes.

Author(s)

Ken Kelley (<kkelley@end.edu>)

References

Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724

Anderson, S. F., & Maxwell, S. E. (2017). Addressing the ‘replication crisis’: Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.

Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

See Also

[ss_buc_independent_t](#) and the other ss_buc_* planners, whose results this function takes as input.

Examples

```
plan <- ss_buc_independent_t(t_observed = 3, n = c(50, 55), assurance = .80)
plan$value[plan$term == "ncp_adjusted"]

# Take the corrected estimate at face value as the truth.
ss_buc_sensitivity(plan, true_ncp = 1.9, replications = 200, seed = 2017)

# And ask what happens if the truth is smaller than the correction assumed.
ss_buc_sensitivity(plan, true_ncp = 1.2, replications = 200, seed = 2017)
```

ss_buc_welch_t	<i>Necessary sample size to reach desired power for a Welch (unequal variance) t test using a publication bias and uncertainty correction procedure</i>
----------------	---

Description

ss_buc_welch_t returns the necessary per-group sample size to achieve a desired level of statistical power for a planned study using a Welch *t* test, the two-group comparison that does not assume equal variances, based on information obtained from a previous study. The effect from the previous study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The bias and uncertainty adjusted previous study noncentrality parameter is also returned, which can be transformed to various effect size metrics.

Usage

```
ss_buc_welch_t(
  t_observed,
  n_1,
  n_2,
  sd_ratio = 1,
  sd_1,
  sd_2,
  var_1,
  var_2,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8
)
```

Arguments

t_observed	Observed Welch t value from a previous study used to plan sample size for a planned study. Either sign is accepted.
n_1, n_2	Group sample sizes of the previous study.
sd_ratio	Ratio of the second group's standard deviation to the first group's, in the previous study, assumed to hold in the planned study as well. A value of 1 means equal variances. Give this, or the two group standard deviations, or the two group variances, but not more than one of the three.
sd_1, sd_2	Group standard deviations of the previous study, an alternative to sd_ratio. Only their ratio enters the computation, so the unit does not matter.
var_1, var_2	Group variances of the previous study, a second alternative to sd_ratio.
alpha_prior	Alpha level α for the previous study or the assumed statistical significance necessary for publishing in the field; to assume no publication bias, a value of 1 can be entered.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance, or the long run proportion of times that the planned study power will reach or surpass the desired level (assurance > .5 corrects for uncertainty; assurance < .5 is not recommended). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as .80); a value of exactly 1 is read as 1 percent. A percentage is echoed in the result's planning-input rows as the coerced proportion (80 is echoed as .80); the same applies to desired_power.
desired_power	Desired level of statistical power for the planned study. Enter it as a proportion in (0, 1) or as a percentage greater than 1; a value of exactly 1 is read as 1 percent.

Details

Researchers often use the sample effect size from a prior study as an estimate of the likely size of an expected future effect in sample size planning. However, sample effect size estimates should not

usually be used at face value to plan sample size, due to both publication bias and uncertainty.

This planner is an approximation, and needs one assumption the others do not. A Welch t statistic is not exactly distributed as a noncentral t : its denominator degrees of freedom are estimated from the data by the Welch-Satterthwaite formula, and they depend on the ratio of the two population standard deviations. `ss_buc_welch_t` treats the statistic as noncentral t on the Welch-Satterthwaite degrees of freedom implied by the prior study's group sizes and the standard deviation ratio you supply, which is the usual approximation in power analysis for this test. In simulation that treatment is accurate: across ratios from 1 to 3 the Type I error of the real Welch test stays between .047 and .054 against a nominal .05, and the planned study's power is within about .006 of what this function predicts, with the agreement no worse at the extreme ratios.

The assumption that matters is the ratio itself. Planning assumes the ratio you supply is the ratio the planned study will have, and getting that wrong costs far more than the distributional approximation does. For the example below, a plan built assuming equal variances delivers about .43 power rather than .80 if the true ratio turns out to be 2. Re-run across a range of plausible ratios and plan for the least favorable one you find credible. The required sample size does not always rise with the ratio: it falls when the group assumed to be more variable is the smaller one.

When the ratio is 1 and the prior groups were the same size, this planner returns exactly what `ss_buc_independent_t` returns. With unequal prior group sizes the two differ, and should: the equal-variance planner reduces the prior study to an equivalent equal- n design at the harmonic mean, while this one uses the actual Welch-Satterthwaite degrees of freedom, which are smaller when the groups are unbalanced. The same prior t is then less compelling and the plan is larger.

The approach uses a likelihood function of a truncated noncentral F distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is the density of a noncentral F distribution. The denominator is the power of the test, which serves to truncate the distribution. In the two-group case, this formula reduces to the density of a truncated noncentral t distribution. (See Taylor & Muller, 1996, Equation 2.1, and Anderson & Maxwell, 2017, for more details.)

Assurance is the proportion of times that power will be at or above the desired level, if the experiment were to be reproduced many times. For example, assurance = .5 means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the (1 - assurance) quantile of the likelihood distribution.

If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), `alpha_prior` can be set to 1 to indicate no publication bias. Alternative α levels can also be accommodated to represent differing amounts of publication bias. For example, setting `alpha_prior` = .20 would reflect less severe publication bias than the default of .05. In essence, setting `alpha_prior` at .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and bias. When this happens, subsequent sample size planning is not possible with the chosen specifications, and the function stops with an informative error rather than returning a result. The error reports the largest assurance the

prior result can support, which is $1 - p/\alpha$, where p is the prior study's p value and α is `alpha_prior`; at any higher assurance the corrected noncentrality parameter is driven to zero. Two remedies are available. First, users can select a lower value of assurance (e.g., .8 instead of .95), at or below the reported ceiling. Second, users can reduce the influence of publication bias by setting `alpha_prior` at a value greater than .05, which raises the ceiling. Note that `alpha_prior` is not a property of the prior study, which is fixed, but the analyst's assumption about the publication threshold its literature faced. When the ceiling is at or below the recommended floor of .5, lowering assurance cannot help and `alpha_prior` is the only remaining lever. It is possible to correct for uncertainty only by setting `alpha_prior` = 1 and choosing the desired level of assurance. We encourage users to make the adjustments as minimal as possible.

The returned `actual_power` is the statistical power of the plan this function reports: the power of a study of the returned sample size when the effect is the returned bias and uncertainty adjusted noncentrality parameter. Because a sample size must be a whole number, `actual_power` is never below the `desired_power` that was requested and is usually a little above it.

The observed t may be entered with either sign; the publication rule the correction assumes is two-sided, so only the magnitude enters the computation. The planned study is assumed to have equal group sizes, which is the allocation that maximizes power when the variances are equal and is a reasonable default otherwise. It is not the optimal allocation when they are not: sampling in proportion to the standard deviations would need about 11 percent fewer participants in total at a ratio of 2, and about 25 percent fewer at a ratio of 3.

Value

An object of class `bucss_power`: a `data.frame` with a character `term` column and a numeric `value` column, in the same shape as the sibling package `DMAR`'s planners. The design results come first: the necessary per-group sample size for the planned study, named for its unit (`necessary_n_per_group`, `necessary_n_per_cell`, or `necessary_N`); for the per-group and per-cell designs, `total_N`, the implied total sample size; `actual_power`, the conservative achieved power at the returned sample size (see Details); and `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter. Rows echoing the planning inputs follow, so the assumptions travel with the result. The design, the unit the sample size is counted in, the largest assurance this prior can support (the assurance ceiling), and the planned test's degrees of freedom (`df_effect` and `df_error`) travel on attributes; they are shown by `print.bucss_power`, and `tidy()` and `glance()` give the compact and wide views.

Author(s)

Ken Kelley (<kkelley@end.edu>)

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29, 844–867. doi:10.1037/met0000520
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547–1562. doi:10.1177/0956797617723724

Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2027). *Designing experiments and analyzing data: A model comparison perspective* (4th ed.). Routledge.

Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25, 1595–1610.

See Also

[ss_buc_independent_t](#) for the equal-variance test.

Examples

```
# A prior study with 40 and 55 participants whose second group was half
# again as variable as the first.
result <- ss_buc_welch_t(t_observed = 3, n_1 = 40, n_2 = 55,
  sd_ratio = 1.5, alpha_prior = .05, alpha_planned = .05,
  assurance = .80, desired_power = .80)
result

# The same study described by the group standard deviations, or by the
# group variances, rather than by their ratio.
ss_buc_welch_t(t_observed = 3, n_1 = 40, n_2 = 55, sd_1 = 8.4, sd_2 = 12.6)
ss_buc_welch_t(t_observed = 3, n_1 = 40, n_2 = 55, var_1 = 70.56,
  var_2 = 158.76)

# How sensitive is the plan to the assumed spread?
ss_buc_welch_t(t_observed = 3, n_1 = 40, n_2 = 55, sd_ratio = 1)
ss_buc_welch_t(t_observed = 3, n_1 = 40, n_2 = 55, sd_ratio = 2)

# Requesting more assurance than the prior result can support stops with an
# informative error naming the largest workable assurance (here near .93):
try(ss_buc_welch_t(t_observed = 3, n_1 = 40, n_2 = 55, sd_ratio = 1.5,
  assurance = .95))
```

Index

BUCSS (BUCSS-package), 2
BUCSS-package, 2
bucss_tidiers, 3

glance.bucss_power (bucss_tidiers), 3

planning_sentence, 5
plot.bucss_power, 5

ss_buc_ancova, 8
ss_buc_chisq_diff, 11
ss_buc_correlation, 15
ss_buc_factorial_anova, 19, 23, 43
ss_buc_factorial_anova_general, 9, 11, 19, 22, 43
ss_buc_independent_t, 26, 68, 70, 72
ss_buc_manova, 30
ss_buc_mixed_anova, 34
ss_buc_mixed_anova_general, 37
ss_buc_one_sample_t, 41
ss_buc_one_way_anova, 9, 19, 23, 43
ss_buc_paired_t, 42, 43, 46
ss_buc_R2, 50
ss_buc_reg_coef, 16–18, 53
ss_buc_reg_joint, 56
ss_buc_rm_anova, 60
ss_buc_rm_anova_general, 63
ss_buc_sensitivity, 7, 66
ss_buc_smd (ss_buc_independent_t), 26
ss_buc_smd_paired (ss_buc_paired_t), 46
ss_buc_welch_t, 67, 68

tidy.bucss_power (bucss_tidiers), 3

uniroot, 3