

Package ‘BUCSS’

October 14, 2026

Type Package

Title Bias and Uncertainty Corrected Sample Size

Version 2.0.0

Date 2026-10-13

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. Planners cover the one-sample, paired, independent, and Welch t tests; between-subjects analysis of variance and covariance; within-subjects and split-plot analysis of variance; a regression coefficient, the squared multiple correlation, and a subset of predictors; the Pearson correlation; a two-group multivariate comparison; and the chi square difference between nested models. The regression planners can treat the predictors as fixed or as sampled, the within-subjects and split-plot planners accept a sphericity correction, and each plan reports the effect size it is built on. A simulation function audits how often a plan attains its target power. The method is described in Anderson, Kelley, and Maxwell (2017) <[doi:10.1177/0956797617723724](https://doi.org/10.1177/0956797617723724)> and extended to replication studies in Anderson and Kelley (2024) <[doi:10.1037/met0000520](https://doi.org/10.1037/met0000520)>.

License GPL (>= 3)

Encoding UTF-8

Language en-US

Depends R (>= 3.6.0)

Imports generics, graphics, stats, withr (>= 3.0.3)

Suggests testthat (>= 3.1.7), 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

Config/roxygen2/version 8.1.0

NeedsCompilation no

Author Ken Kelley [aut, cre] (ORCID: <<https://orcid.org/0000-0002-4756-8360>>),
Samantha F. Anderson [aut] (ORCID:
<<https://orcid.org/0000-0003-4451-5295>>),
Scott E. Maxwell [aut] (ORCID: <<https://orcid.org/0000-0001-8067-7027>>)

Maintainer Ken Kelley <kkelley@nd.edu>

Repository CRAN

Date/Publication 2026-10-13 23:00:30 UTC

Contents

BUCSS-package	3
bucss_tidiers	5
planning_sentence	7
plot.bucss_plan	8
ss_buc_ancova	10
ss_buc_chisq_diff	15
ss_buc_correlation	20
ss_buc_factorial_anova	25
ss_buc_factorial_anova_general	30
ss_buc_independent_t	35
ss_buc_manova	40
ss_buc_mixed_anova	44
ss_buc_mixed_anova_general	51
ss_buc_one_sample_t	57
ss_buc_one_way_anova	61
ss_buc_paired_t	65
ss_buc_R2	70
ss_buc_reg_coef	75
ss_buc_reg_joint	81
ss_buc_rm_anova	86
ss_buc_rm_anova_general	92
ss_buc_sensitivity	98
ss_buc_welch_t	101

Index

107

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

Each `ss_buc_*` function plans one analysis. The *t* tests: `ss_buc_one_sample_t`, `ss_buc_paired_t`, `ss_buc_independent_t`, and `ss_buc_welch_t`. Between-subjects analysis of variance and covariance: `ss_buc_one_way_anova`, `ss_buc_factorial_anova`, `ss_buc_factorial_anova_general`, and `ss_buc_ancova`. Within-subjects and split-plot analysis of variance: `ss_buc_rm_anova`, `ss_buc_rm_anova_general`, `ss_buc_mixed_anova`, and `ss_buc_mixed_anova_general`. Regression and correlation: `ss_buc_reg_coef`, `ss_buc_R2`, `ss_buc_reg_joint`, and `ss_buc_correlation`. And `ss_buc_manova` for a two-group multivariate comparison and `ss_buc_chisq_diff` for the chi square difference between nested models. The experimental designs among these, their tests, and the statistical power of those tests are described in Maxwell, Delaney, and Kelley (2027). `ss_buc_sensitivity` simulates the literature a plan came from and reports how often such a plan attains its target power, and `plot` draws the largest assurance a prior result can support.

The planners carried over from BUCSS 1.x are also available, through a web interface, as Shiny apps at <https://designingexperiments.com>; the designs added in 2.0.0 are available in R only.

BUCSS follows the design and naming conventions of the **DMAR** package (Design, Measurement, and Analysis in R): 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.

How assurance is used. Every planner takes an assurance: the long run proportion of planned studies whose power at the true effect reaches or surpasses the desired power, when each is planned from a prior study published under the rule `alpha_prior` describes. A planned study reaches its target power whenever the noncentrality parameter it was planned from is at or below the true one, because power rises with the effect (and, apart from rounding the sample size up to a whole number and from the smallest sample size a planner returns, the design's own minimum or the Welch planner's 10 per group, only then). The planners therefore plan from the lower limit of a one-sided confidence interval for the prior study's noncentrality parameter, at confidence level assurance, computed from the distribution of the prior statistic truncated at the publication threshold: the value at which a published prior study would exceed the observed statistic with probability $1 - \text{assurance}$ (what Anderson, Kelley, and Maxwell, 2017, call the $(1 - \text{assurance})$ quantile of the likelihood distribution). At assurance = .5 this is the median unbiased estimate, which corrects for publication

bias only; above .5 it adds the correction for uncertainty. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired; here it is the probability that the planned study's power will be at least the desired power. `ss_buc_sensitivity` checks the proportion by simulation for any plan, counting a refusal to plan as meeting the target, as the guarantee does, and also reporting the rate among the plans issued.

The planning effect size. Every plan also reports a planning effect size, in a row whose name ends in `_planning` (`smd_planning`, `f_planning`, `rho_planning`, and so on). It is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through assurance). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Each planner's Value section names its metric.

Safeguard power is a special case. Safeguard power (Perugini, Gallucci, & Costantini, 2014) is a way of planning a new study from a prior study's effect size that guards against the prior estimate being too large. Instead of planning for the effect size the prior study observed, the researcher plans for the lower limit of a two-sided 60 percent confidence interval around it (a limit that, in repeated sampling, falls below the true effect 80 percent of the time), so the new study keeps its intended power even if the true effect is somewhat smaller than the one observed. It allows for uncertainty in the prior estimate but not for publication bias.

It is the special case `alpha_prior = 1` and `assurance = .80` of the method here, that is, BUCSS without the publication bias correction, at an assurance of .80. `alpha_prior` is the significance level a prior result had to reach to be published, and no p value exceeds 1, so `alpha_prior = 1`, which may also be written `alpha_prior = "No Publication Filter"`, means no publication filter: any result would have been published, the truncation is removed, and the prior statistic has its ordinary distribution. (Anderson, Kelley, and Maxwell, 2017, p. 1552, note that this level can be set equal to 1 when publication bias is thought not to apply.) At `assurance = .80` the adjusted noncentrality parameter is then the lower limit of a one-sided 80 percent confidence interval, which is the lower limit of the two-sided 60 percent interval (either way, 20 percent lies below it), and the plan is the smallest sample size whose ordinary power at that value reaches `desired_power`. In general, the lower limit of a two-sided interval at confidence level C is `assurance = (1 + C)/2`, so a two-sided 80 percent interval is `assurance = .90`.

For an F or chi square statistic the correspondence is exact: the adjusted noncentrality parameter is the lower limit of that interval, computed from the statistic's ordinary noncentral distribution. Two differences from safeguard power remain, and each can only make the plan here the same as the safeguard plan or larger. First, for the t tests and every other effect on one degree of freedom, safeguard power computes its interval from the signed t , while the interval here bounds the size of the effect whatever its sign, because the publication rule is two-sided. That difference is negligible for a convincing prior result and grows as the prior result weakens. Second, an unbalanced prior design is read as a balanced one, and the Welch planner returns at least 10 per group. `vignette("understanding-assurance", package = "BUCSS")` measures both. For the regression planners the equivalence holds in the frame `random_predictors` names. With `alpha_prior` below 1 the same construction is applied to the publication-truncated distribution, which is what the bias correction adds to safeguard power.

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

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

Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305

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

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science*, 9, 319–332. doi:10.1177/1745691614528519

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. doi:10.1080/03610929608831787

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, as the sibling package DMAR’s sample size planners do: one row with term ("sample_size"), estimate (the necessary sample size in the design’s unit, which the result’s sample_size_unit attribute names), and power (the power of the plan the result reports; see actual_power on the planner’s help page). glance() returns the one-row wide view: the same three columns, then one column per row of the result, namely the sample size row, total_N where applicable, actual_power, ncp_adjusted, the planning effect size (the row whose name ends in _planning, such as smd_planning), and every echoed planning input, together with the design, the effect tested, the unit, the planned test’s degrees of freedom where the design has them, and the assurance ceiling.

Usage

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

Arguments

x	An object of class <code>bucss_plan</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. Both stop with an explanation if the object no longer has the shape of a single plan (for example, several plans stacked with `dplyr::bind_rows()`, or a plan with a row removed by `dplyr::filter()`; both keep the class).

Author(s)

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

Description

Turns a `bucss_plan` 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 stating which corrections the plan made. The clause is built from the plan's own settings: the publication bias correction is named, with its assumed threshold, only when `alpha_prior` is below 1, and the uncertainty correction only when assurance is above .5 (at .5 the plan uses the median estimate). A within-subjects or split-plot plan made with `epsilon` below 1 adds the departure from sphericity it allowed for. A second sentence names the planning effect size (the result's row whose name ends in `_planning`), the adjusted effect size the plan is built on, in the design's usual metric: a standardized mean difference, Cohen's f , a correlation, and so on, as the planner's help page describes. It is the value used to plan given the prior result, `alpha_prior`, and assurance, not an estimate of the effect to report. It is shown at `getOption("bucss.digits", 3)` significant figures (more when fewer would show a value below 1 as 1), and without a leading zero when the metric cannot exceed 1. This mirrors the `results_sentence()` helper in the sibling package DMAR.

Usage

```
planning_sentence(x)
```

Arguments

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

Value

A single character string holding the two sentences.

Author(s)

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

Examples

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

# With no publication bias assumed (a pilot study, say), the sentence
# names the uncertainty correction alone.
planning_sentence(ss_buc_independent_t(t_observed = 3, n = 20,
  alpha_prior = 1))
```

plot.bucss_plan	<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_plan'
plot(
  x,
  which = c("both", "ncp", "size"),
  assurance = NULL,
  lower = NULL,
  n_points = 60L,
  col = "#B03A2E",
  ...
)
```

Arguments

<code>x</code>	A plan returned by any <code>ss_buc_*</code> function.
<code>which</code>	Which panel to draw: "both" (the default, side by side), "ncp" for the corrected noncentrality parameter alone, or "size" for the sample size alone.
<code>assurance</code>	Assurance values to evaluate. NULL (the default) builds a grid running from lower up to just short of the ceiling.
<code>lower</code>	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.
<code>n_points</code>	Number of assurance values in the automatic grid.
<code>col</code>	Color for the curve and the marked plan.
<code>...</code>	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, where a test's power is its Type I error rate whatever the sample size (on statistical power, see Maxwell, Delaney, & Kelley, 2027), so no sample size attains the target power, and the planner stops. The ceiling is reported in the printed footer of every plan and, when it is above .5, named in that error, but its shape is the point: assurance does not degrade gracefully

as it rises, it hits a wall. For a planner that reduces a prior study with unequal group sizes, or a total sample size that does not divide evenly, to a balanced design, the largest supportable assurance is computed at that design, so it can be below $1 - p/\alpha$ at the study's own p value.

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 when it lies in the plotted range (a plan made at an assurance below lower is marked once lower is set below it), 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>) 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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305

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. doi:10.1080/03610929608831787

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, together with the planning effect size, the effect size the plan is built on, as Cohen’s *f*.

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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an alpha_prior of 1.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses desired_power, when each is planned from a prior study published under the rule alpha_prior describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so assurance = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With alpha_prior = 1 and assurance = .80 the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, &

Kelley, 2027). It must exceed `alpha_planned`: a test's power is never below its Type I error rate, so a value at or below `alpha_planned` is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless `alpha_planned` is below .01.

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 (Maxwell, Delaney, & Kelley, 2027), 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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

`alpha_prior` is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below `alpha_prior` reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting `alpha_prior = .20` assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set `alpha_prior = 1`, which may also be written `alpha_prior = "No Publication Filter"`; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by `assurance`, remains.

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, chance imbalance between the groups on the covariates lowers the planned study's power below the reported value, by roughly .4 to .5 times q/N at a desired power of .80 and an `alpha_planned` of .05 (somewhat more at a stricter `alpha_planned`, or when the plan has few participants per covariate), for q covariates and N planned participants in total: about .002 for the first example below (462 participants, two covariates), but about .05 for a plan of 27 participants with three covariates.

`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. The planning effect size belongs to the same balanced design as the reported noncentrality parameter, so an ordinary power analysis at it can return a sample size a little smaller than the one reported here (never a larger one), because the reported size is the larger of the two rounded designs' requirements. For a prior study with unequal group sizes, or a total sample size that does not divide evenly, the largest supportable assurance is computed at the balanced design the planner plans from, so it can be below $1 - p/\alpha$ at the study's own p value.

Value

An object of class `bucss_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic, at the prior study's sample size (for a prior N that does not divide evenly, at the balanced sample size the planner substitutes for it; see Details); and the planning effect size, `f_planning`, Cohen's f for the effect tested, relative to the covariate-adjusted error variance, $\sqrt{\lambda/N}$, where λ is `ncp_adjusted` and N the total of the balanced prior design it belongs to (see Details), so $\lambda = Nf^2$, the convention

of power analysis for between-subjects designs, and $f^2/(1+f^2)$ is the population partial eta squared for the effect.

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through `assurance`). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787

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, together with the planning effect size, the effect size the plan is built on, as the root mean square error of approximation of the difference test, $RMSEA_D$ (D for difference).

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, the signature of an omnibus test of model fit rather than a difference test between two substantive models; the function refuses that case (see Details, which also covers a just-identified full model).
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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an alpha_prior of 1.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses desired_power, when each is planned from a prior study published under the rule alpha_prior describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so assurance = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With alpha_prior = 1 and assurance = .80 the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed alpha_planned: a test's power is never below its Type I error rate, so a value at or below alpha_planned is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless alpha_planned is below .01.

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. Steiger, Shapiro, and Browne (1985) give the asymptotic distribution of sequential chi square statistics for nested models and of the difference tests between them, and Satorra and Saris (1985) compute the power of the likelihood ratio test in covariance structure analysis from the noncentral chi square distribution, as is done here. 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 $df_{full} = 0$ is its signature. 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. A just-identified full model (every parameter free, as in a path model compared with the same model with one path fixed at zero) also has zero degrees of freedom, and there the difference test is a genuine one. The plan depends only on the number of constraints, so for that case the call with $df_{full} = 1$ and $df_{restricted}$ one more than the number of constraints gives the plan.

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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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 $(1 - assurance)$ quantile of the likelihood distribution.

$alpha_prior$ is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below $alpha_prior$ reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting $alpha_prior$

$= .20$ assumes that studies with p values less than $.20$ are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of $.05$. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set `alpha_prior` = 1, which may also be written `alpha_prior = "No Publication Filter"`; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by assurance, remains.

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 likelihood ratio test is liberal in small samples, and more so the more parameters a model estimates. For a test that drops 2 of 6 parameters of a normal linear model, its actual Type I error is about $.07$ at a total sample size of 50, $.055$ at 200, and $.052$ at 400; dropping 5 of 21 parameters, it is about $.08$ at 200. At such sample sizes the real test's power can also sit above or below `actual_power`. Treat a planned sample size that is small relative to the number of parameters estimated with caution, and read a very small one as a signal that the prior effect was large rather than as a serious recommendation. The search for the planned sample size starts at the number of constraints plus 2 (`df_restricted` - `df_full` + 2), so no plan is smaller than that, even when a smaller study would already reach the desired power.

Value

An object of class `bucss_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see `Details`); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the chi square statistic, at the prior study's sample size; and the planning effect size, `rmsea_d_planning`,

RMSEA_D, where the D stands for difference: the root mean square error of approximation of the chi square difference test between the two nested models (Savalei, Brace, & Fouladi, 2024), as distinct from the RMSEA of either model's fit; here $\sqrt{\lambda/(df_D N)}$, where λ is ncp_adjusted, df_D is the number of constraints ($df_{\text{restricted}} - df_{\text{full}}$, the difference test's degrees of freedom), and N the prior sample size, the scaling this planner plans with; the usual sample formula divides by $N - 1$ instead, which gives a value larger by the factor $\sqrt{N/(N - 1)}$.

A planning effect size (the row whose name ends in _planning) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through alpha_prior, and the uncertainty allowed for through assurance). It is ncp_adjusted restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates ncp_adjusted, actual_power is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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 where the test has a denominator) travel on attributes. print.bucss_plan shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; tidy() and glance() give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787
- Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, 50, 83–90. doi:10.1007/BF02294150

Savalei, V., Brace, J. C., & Fouladi, R. T. (2024). We need to change how we compute RMSEA for nested model comparisons in structural equation modeling. *Psychological Methods*, 29, 480–493. doi:10.1037/met0000537

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, 50, 253–263. doi:10.1007/BF02294104

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, together with the planning effect size, the effect size the plan is built on, as a correlation.

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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an alpha_prior of 1.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses desired_power, when each is planned from a prior study published under the rule alpha_prior describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so assurance = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With alpha_prior = 1 and assurance = .80 the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed alpha_planned: a test's power is never below its Type I error rate, so a value at or below alpha_planned is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless alpha_planned is below .01.

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; Anderson, Kelley, & Maxwell, 2017, Equation 3; and Anderson & Maxwell, 2017, for more details.) Which density belongs in that numerator is described next; it is the density the regression planners use with `random_predictors = TRUE`.

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 : the estimates, and the distribution of the statistic when there is no effect, are the same whether the predictor is fixed or sampled, but the distribution when there is an effect, and so the power, is not (Sampson, 1974; Gatsonis & Sampson, 1989; Kelley & Maxwell, 2003). 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 in every case, by roughly $1/n$ for a plan of n participants: about .01 at $n = 100$, .02 at 50, .05 at 20, .10 at 10, and up to about .2 for the smallest plans.

`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. Stated for the correlation itself, it is the general sampling distribution of the multiple correlation coefficient (Fisher, 1928) with one predictor. 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 (1 to 3 for the smallest plans, up to about 8 for plans in the hundreds). Near the assurance ceiling the difference reverses and the mixture calls for a sample a few percent smaller, which can be hundreds of participants because plans there are very large (7,346 rather than the 7,722 of treating the predictor as fixed, for $r = .50$ from a prior study of 20 at an assurance of .5). 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` and `random_predictors = FALSE` is the planner for it. With one sampled predictor the regression planners reduce to this one: `ss_buc_reg_coef(p = 1, random_predictors = TRUE)` gives this planner's own result for the same prior t and N , exactly, as do `ss_buc_R2` with `p = 1` and `ss_buc_reg_joint` with `p = 1` and `p_joint = 1` given `F_observed = t^2`; each states the planning effect size in its own metric.

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.

alpha_prior is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below alpha_prior reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting alpha_prior = .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set alpha_prior = 1, which may also be written alpha_prior = "No Publication Filter"; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by assurance, remains.

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 planning effect size implies at the prior study's sample size, $N\rho^2/(1 - \rho^2)$, on the scale of the F statistic, as in the regression planners (the t -based planners report theirs on the scale of t instead). The planning effect size itself, the correlation $\sqrt{\text{ncp_adjusted} / (\text{ncp_adjusted} + N)}$, is returned as the rho_planning row.

Value

An object of class `bucss_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic, $N\rho^2/(1 - \rho^2)$ at the prior study's sample size; and the planning effect size, `rho_planning`, the population correlation (a magnitude), $\sqrt{\lambda/(\lambda + N)}$, where λ is `ncp_adjusted` and N the prior sample size; the power statement that follows is for sampled variables, the exact distribution this planner uses.

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through assurance). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787

Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 121, 654–673. doi:10.1098/rspa.1928.0224

Gatsonis, C., & Sampson, A. R. (1989). Multiple correlation: Exact power and sample size calculations. *Psychological Bulletin*, 106, 516–524. doi:10.1037/00332909.106.3.516

Sampson, A. R. (1974). A tale of two regressions. *Journal of the American Statistical Association*, 69, 682–689. doi:10.1080/01621459.1974.10480189

See Also

[ss_buc_reg_coef](#) for a coefficient in a model with more than one predictor (with random_predictors = TRUE for sampled predictors).

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, together with the planning effect size, the effect size the plan is built on, as Cohen's f .

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

<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>levels_A</code>	Number of levels for Factor A.
<code>levels_B</code>	Number of levels for Factor B.
<code>effect</code>	Effect of most interest to the planned study: main effect of A (<code>factor_A</code>), main effect of B (<code>factor_B</code>), or interaction (<code>interaction</code>). Case-insensitive shorthand is also accepted but not recommended in scripts: "A" (or "a") for <code>factor_A</code> , "B" for <code>factor_B</code> , and "AxB" (or "AB") for interaction.
<code>alpha_prior</code>	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below <code>alpha_prior</code> . Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by <code>assurance</code>) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an <code>alpha_prior</code> of 1.
<code>alpha_planned</code>	Alpha level α assumed for the planned study.
<code>assurance</code>	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses <code>desired_power</code> , when each is planned from a prior study published under the rule <code>alpha_prior</code> describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so <code>assurance</code> = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and

Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With `alpha_prior = 1` and `assurance = .80` the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see [BUCSS-package](#) for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed `alpha_planned`: a test's power is never below its Type I error rate, so a value at or below `alpha_planned` is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless `alpha_planned` is below .01.

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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

`alpha_prior` is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below `alpha_prior` reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting `alpha_prior = .20` assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set `alpha_prior`

= 1, which may also be written `alpha_prior = "No Publication Filter"`; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by assurance, remains.

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. The planning effect size belongs to the same balanced design as the reported noncentrality parameter, so an ordinary power analysis at it can return a sample size a little smaller than the one reported here (never a larger one), because the reported size is the larger of the two rounded designs' requirements. For a prior study with unequal group sizes, or a total sample size that does not divide evenly, the largest supportable assurance is computed at the balanced design the planner plans from, so it can be below $1 - p/\alpha$ at the study's own p value. 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_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic, at

the prior study's sample size (for a prior N that does not divide evenly, at the balanced sample size the planner substitutes for it; see Details); and the planning effect size, f_{planning} , Cohen's f for the effect tested, relative to the error term of its test, $\sqrt{\lambda/N}$, where λ is npc_adjusted and N the total of the balanced prior design it belongs to (see Details), so $\lambda = Nf^2$, the convention of power analysis for between-subjects designs, and $f^2/(1 + f^2)$ is the population partial eta squared for the effect.

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through `assurance`). It is npc_adjusted restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates npc_adjusted , `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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. `print.bucss_plan` shows the design, the effect, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787

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, together with the planning effect size, the effect size the plan is built on, as Cohen's f .

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

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.
cells	Number of cells for the design (the product of the number of levels of each factor).
df_numerator	Numerator degrees of freedom for the effect of interest.
df_denominator	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 $N - \text{cells}$.
alpha_prior	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below <code>alpha_prior</code> . Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an <code>alpha_prior</code> of 1.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses <code>desired_power</code> , when each is planned from a prior study published under the rule <code>alpha_prior</code> describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so <code>assurance = .5</code> plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With <code>alpha_prior = 1</code> and <code>assurance = .80</code> the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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 <code>desired_power</code> .
desired_power	Desired level of statistical power for the planned study: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed <code>alpha_planned</code> : a test's power is never below its Type I error rate, so a value at or below <code>alpha_planned</code> is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless <code>alpha_planned</code> is below .01.

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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

alpha_prior is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below alpha_prior reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting $\text{alpha_prior} = .20$ assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set $\text{alpha_prior} = 1$, which may also be written $\text{alpha_prior} = \text{"No Publication Filter"}$; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by assurance , remains.

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. The planning effect size belongs to the same balanced design as the reported noncentrality parameter, so an ordinary power analysis at it can return a sample size a little smaller than the one reported here (never a larger one), because the reported size is the larger of the two rounded designs' requirements. For a prior study with unequal group sizes, or a total sample size that does not divide evenly, the largest supportable assurance is computed at the balanced design the planner plans from, so it can be below $1 - p/\alpha$ at the study's own p value. 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_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic, at the prior study's sample size (for a prior N that does not divide evenly, at the balanced sample size the planner substitutes for it; see Details); and the planning effect size, `f_planning`, Cohen's f for the effect tested, relative to the error term of its test (after the nuisance parameters `df_denominator` implies), $\sqrt{\lambda/N}$, where λ is `ncp_adjusted` and N the total of the balanced prior design it belongs to (see Details), so $\lambda = Nf^2$, the convention of power analysis for between-subjects designs, and $f^2/(1 + f^2)$ is the population partial eta squared for the effect.

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand

(the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through `assurance`). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an `assurance` of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787

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, together with the planning effect size, the effect size the plan is built on, as a standardized mean difference.

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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an alpha_prior of 1.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses desired_power, when each is planned from a prior study published under the rule alpha_prior describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so assurance = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With alpha_prior = 1 and assurance = .80 the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed alpha_planned: a test's power is never below its Type I error rate, so a value at or below alpha_planned is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless alpha_planned is below .01.

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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

`alpha_prior` is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below `alpha_prior` reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting `alpha_prior` = .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set `alpha_prior` = 1, which may also be written `alpha_prior` = "No Publication Filter"; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by assurance, remains.

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. The planning effect size belongs to the same balanced design as the reported noncentrality parameter, so an ordinary power analysis at it can return a sample size a little smaller than the one reported here (never a larger one), because the reported size is the larger of the two rounded designs' requirements. For a prior study with unequal group sizes, or a total sample size that does not divide evenly, the largest supportable assurance is computed at the balanced design the planner plans from, so it can be below $1 - p/\alpha$ at the study's own p value. 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_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the t statistic (the noncentrality parameter of the prior study's noncentral t); and the planning effect size, `smd_planning`, a standardized mean difference (Cohen's δ , a magnitude), $\lambda\sqrt{2/n}$, where λ is `ncp_adjusted` and n the per-group size of the balanced prior design it belongs to (see Details).

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through assurance). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787

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)
```

```
# A prior study that faced no publication filter (a pilot study, say):
# alpha_prior = "No Publication Filter" means the same as alpha_prior = 1
ss_buc_independent_t(t_observed = 3, n = 20,
  alpha_prior = "No Publication Filter")

# 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, together with the planning effect size, the effect size the plan is built on, as a squared Mahalanobis distance.

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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an alpha_prior of 1.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses desired_power, when each is planned from a prior study published under the rule alpha_prior describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so assurance = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With alpha_prior = 1 and assurance = .80 the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed alpha_planned: a test's power is never below its Type I error rate, so a value at or below alpha_planned is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless alpha_planned is below .01.

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 noncentral 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 (the generalization of Student's ratio; Hotelling, 1931) 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, so the power of the usual tests depends on more than one parameter and this correction, which solves for one, 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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

alpha_prior is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below alpha_prior reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting alpha_prior = .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set alpha_prior = 1, which may also be written alpha_prior = "No Publication Filter"; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by assurance, remains.

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. The planning effect size belongs to the same balanced design as the reported noncentrality parameter, so an ordinary power analysis at it can return a sample size a little smaller than the one reported here (never a larger one), because the reported size is the larger of the two rounded designs' requirements. For a prior study with unequal group sizes, or a total sample size that does not divide evenly, the largest supportable assurance is computed at the balanced design the planner plans from, so it can be below $1 - p/\alpha$ at the study's own p value.

Value

An object of class `bucss_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic, at the prior study's sample size (for an odd prior N , at the balanced sample size the planner substitutes for it; with one outcome variable it is the square of the independent t planner's); and the planning effect size, `D2_planning`, the squared Mahalanobis distance between the two population mean vectors, $2\lambda/n$, where λ is `ncp_adjusted` and n the per-group size of the balanced prior design it belongs to (see Details); with one outcome variable it is the square of the `smd_planning` of [ss_buc_independent_t](#).

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through assurance). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787
- Hotelling, H. (1931). The generalization of Student's ratio. *The Annals of Mathematical Statistics*, 2, 360–378. doi:10.1214/aoms/1177732979

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, together with the planning effect size, the effect size the plan is built on, as Cohen's f .

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,
  epsilon = 1
)
```

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>levels_between</code>	Number of levels for the between-subjects factor.
<code>levels_within</code>	Number of levels for the within-subjects factor.
<code>effect</code>	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.
<code>alpha_prior</code>	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below <code>alpha_prior</code> . Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by <code>assurance</code>) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an <code>alpha_prior</code> of 1.
<code>alpha_planned</code>	Alpha level α assumed for the planned study.
<code>assurance</code>	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses <code>desired_power</code> , when each is planned from a prior study published under the rule <code>alpha_prior</code> describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence

interval at this level, computed from the publication-truncated distribution of the prior statistic, so $\text{assurance} = .5$ plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With $\text{alpha_prior} = 1$ and $\text{assurance} = .80$ the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see [BUCSS-package](#) for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed `alpha_planned`: a test's power is never below its Type I error rate, so a value at or below `alpha_planned` is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless `alpha_planned` is below .01.

`epsilon` The sphericity parameter ϵ of the within-subjects contrasts, for the within-subjects and interaction effects, from $1/(\text{levels_within} - 1)$ (its lower bound, for both effects) to 1 (the default, which assumes sphericity); with two within-subjects levels it takes only 1. For a prior study reported with Greenhouse-Geisser or Huynh-Feldt corrected degrees of freedom, it is the reported numerator degrees of freedom divided by the uncorrected ones. The same value describes the planned study, and it is echoed as the `epsilon` row of the result. It does not apply to the between-subjects effect, whose test is on subject means: there a value other than 1 is ignored with a warning, and the result has no `epsilon` row. See Details.

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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

alpha_prior is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below alpha_prior reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting alpha_prior = .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set alpha_prior = 1, which may also be written alpha_prior = "No Publication Filter"; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by assurance, remains.

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. The planning effect size belongs to the same balanced design as the reported noncentrality parameter, so an ordinary power analysis at it can return a sample size a little smaller than the one reported

here (never a larger one), because the reported size is the larger of the two rounded designs' requirements. For a prior study with unequal group sizes, or a total sample size that does not divide evenly, the largest supportable assurance is computed at the balanced design the planner plans from, so it can be below $1 - p/\alpha$ at the study's own p value. Although equal- n previous studies are preferable, this approach will work well as long as the cell sizes are only slightly discrepant.

Sphericity. The default $\epsilon = 1$ assumes sphericity, as the uncorrected within-subjects test does. When a prior study reports Greenhouse-Geisser (1959) or Huynh-Feldt (1976) corrected degrees of freedom, supply ϵ , the reported numerator degrees of freedom divided by the uncorrected ones. For example, $F(1.6, 92.8)$ for the within-subjects effect of a 2×3 design with 60 subjects gives $\epsilon = 1.6/2$, and the denominator agrees: $92.8/116$ is also $.8$. The ratio of the two reported degrees of freedom, $92.8/1.6 = 58$, is the number of subjects the analysis used less the number of groups (for the interaction, that difference divided by one less than the number of groups); when it points to a different number of subjects than the study reports, that number is the N to supply. An implied ϵ above 1 or below its lower bound (see Arguments) means the reported degrees of freedom are not a Greenhouse-Geisser or Huynh-Feldt correction of this effect; a Huynh-Feldt estimate above 1 is reported as the uncorrected degrees of freedom, which is $\epsilon = 1$. `ss_buc_mixed_anova` then applies the approximation of Muller and Barton (1989) to both studies: the prior F is treated as noncentral F with both degrees of freedom and the noncentrality parameter multiplied by ϵ , the publication threshold is the critical value at the corrected degrees of freedom, and the planned study's power is computed the same way. It applies to the within-subjects effect and the interaction, which share the within-subjects contrasts and so share one ϵ ; the between-subjects test is on subject means and does not involve sphericity. The returned `ncp_adjusted` stays on the sphericity scale, so it is comparable across values of ϵ , and the planned test's corrected degrees of freedom are the `df_effect` and `df_error` attributes. One ϵ describes both studies, because it is a property of the covariance matrix of the repeated measures, which the planned study shares with the prior one; the reported value is the prior study's estimate of it, and re-running across the values you find plausible shows how much the plan depends on it. The value used is echoed as the ϵ row of the result, 1 when sphericity is assumed.

The correction under ϵ is an approximation. Below sphericity the distribution of a within-subjects F depends on the whole covariance matrix of the repeated measures, and on how the effect lines up with it, not on ϵ alone. In simulations of raw data (one-way designs of 4 and 6 levels with ϵ from $.45$ to $.86$, prior studies of 12 to 40 subjects published when their Greenhouse-Geisser test reached $p < .05$, and a 2×4 split-plot interaction), planning from each published study with its own reported Greenhouse-Geisser estimate at an assurance of $.80$ gave a plan reaching $.80$ power for the Greenhouse-Geisser test in $.70$ to $.80$ of replications when the effect followed the dominant pattern of variation (a linear trend over autocorrelated occasions), and in $.92$ to $.99$ when it lay in the least variable contrasts. Ignoring the departure ($\epsilon = 1$) reached it in $.26$ to $.64$ of replications for the trend-like effects and in $.70$ to $.87$ for the effects in the least variable contrasts. The shortfall does not shrink as the prior study grows, and is not larger near the lower bound of ϵ . The planned study's power is approximated the same way. Separate simulations of planned one-way studies of 10 to 30 subjects, with 4, 6, or 8 levels, measured the real power of the Greenhouse-Geisser test, whose estimate of ϵ is biased downward. For a trend-like effect (a linear trend over occasions with first-order autoregressive correlations of $.3$ to $.9$, ϵ from $.50$ to $.94$) it fell below `actual_power` by about $.04$ to $.08$ with four levels and $.06$ to $.13$ with six, the larger shortfalls for the smaller plans; a planned study analyzed with the Huynh-Feldt correction tracks `actual_power` more closely. Even when sphericity holds and $\epsilon = 1$, a planned study analyzed with the Greenhouse-Geisser correction anyway, as some software does by default, has less power than `actual_power`, the power of the uncorrected test: about $.02$ to $.06$ less with four levels, $.03$ to $.11$ with six, and $.04$ to $.15$ with

eight, for plans of 30 down to 10 subjects (the Huynh-Feldt test is within about .02). When a paper reports both corrections, supply the Greenhouse-Geisser one: the Huynh-Feldt estimate is larger, and planning from it attained about .02 to .09 less, the larger differences from the smaller prior studies. When sphericity in fact holds but the prior study reported a Greenhouse-Geisser correction anyway, supplying it is conservative, with a median plan 7 to 34 percent larger.

The planning effect size and other power software. Here f_{planning} is Cohen's f relative to the effect's own error term, whose variance already reflects the correlation among the repeated measures, so $f^2/(1 + f^2)$ is the population value of the partial eta squared a paper reports for the effect. It is not the f that power functions for repeated measures designs usually take, such as `ss_power_split_plot_anova()` in the sibling package DMAR: those standardize the effect by the standard deviation of the raw scores and take the correlation among the measures as a separate argument, so giving them f_{planning} , with any correlation, plans a different study. The value that reproduces this plan in `ss_power_split_plot_anova()`, for any of the three effects, with the correlation set to 0 and the same epsilon, is f_{planning} times $\sqrt{k/b}$, where b is `levels_within` and k is the error degrees of freedom per subject (1 for the between-subjects effect, $b - 1$ for the other two). The between-subjects effect is included: its f_{planning} is relative to the variance of subject means, not of raw scores.

Value

An object of class `bucss_plan`: 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 (per between-subjects cell; the row is `necessary_n_per_group`) 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic, at the prior study's sample size (for a prior N that does not divide evenly, at the balanced sample size the planner substitutes for it), and on the sphericity scale for every epsilon (see Details); and the planning effect size, f_{planning} , Cohen's f for the effect tested, relative to its own error term, $\sqrt{\lambda/(kN)}$, where λ is `ncp_adjusted`, N the total of the balanced prior design it belongs to, and k the error degrees of freedom per subject, 1 for the between-subjects effect and `levels_within - 1` for the within-subjects effect and the interaction, so $f^2/(1 + f^2)$ is the population partial eta squared for the effect; like `ncp_adjusted` it is on the sphericity scale, and for epsilon below 1 the power statement that follows holds under the same Muller and Barton (1989) approximation, with the corrected degrees of freedom and noncentrality parameter; it is not the f that power functions for repeated measures take (see Details).

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through assurance). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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. `print.bucss_plan` shows the design, the effect, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112. doi:10.1007/BF02289823
- Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, 1, 69–82. doi:10.3102/10769986001001069
- Muller, K. E., & Barton, C. N. (1989). Approximate power for repeated-measures ANOVA lacking sphericity. *Journal of the American Statistical Association*, 84, 549–555. doi:10.1080/01621459.1989.10478802

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

# A prior study reported with Greenhouse-Geisser corrected degrees of
# freedom of 1.6 and 92.8, a correction of 1.6/2: its plan is larger than
# the one that assumes sphericity (60 per group).
ss_buc_mixed_anova(F_observed = 7, N = 60, levels_between = 2,
  levels_within = 3, effect = "within", epsilon = .8)
```

```
# 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, together with the planning effect size, the effect size the plan is built on, as Cohen's f .

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,
  epsilon = 1
)
```

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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an alpha_prior of 1.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses desired_power, when each is planned from a prior study published under the rule alpha_prior describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so assurance = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With alpha_prior = 1 and assurance = .80 the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed alpha_planned: a test's power is never below its Type I error rate, so a value at or below alpha_planned is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless alpha_planned is below .01.
epsilon	The sphericity parameter ϵ of the within-subjects contrasts, from 1/df_numerator (for within_only) or 1/df_num_within (for between_within), its lower bound, to 1 (the default, which assumes sphericity); when that bound is 1 it takes only 1. For a prior study reported with Greenhouse-Geisser or Huynh-Feldt corrected degrees of freedom, it is the reported numerator degrees of freedom divided by

the uncorrected ones. The same value describes the planned study, and it is echoed as the epsilon row of the result. It does not apply to `between_only`, whose test is on subject means: there a value other than 1 is ignored with a warning, and the result has no epsilon row. See Details.

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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

`alpha_prior` is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below `alpha_prior` reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting `alpha_prior = .20` assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set `alpha_prior = 1`, which may also be written `alpha_prior = "No Publication Filter"`; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by `assurance`, remains.

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. The planning effect size belongs to the same balanced design as the reported noncentrality parameter, so an ordinary power analysis at it can return a sample size a little smaller than the one reported here (never a larger one), because the reported size is the larger of the two rounded designs' requirements. For a prior study with unequal group sizes, or a total sample size that does not divide evenly, the largest supportable assurance is computed at the balanced design the planner plans from, so it can be below $1 - p/\alpha$ at the study's own p value. Although equal- n previous studies are preferable, this approach will work well as long as the cell sizes are only slightly discrepant.

Sphericity. The default `epsilon` = 1 assumes sphericity, as the uncorrected within-subjects test does. When a prior study reports Greenhouse-Geisser (1959) or Huynh-Feldt (1976) corrected degrees of freedom, supply `epsilon`, the reported numerator degrees of freedom divided by the uncorrected ones. For example, $F(1.5, 130.5)$ for a `within_only` effect with 2 numerator degrees of freedom in a study of 90 subjects in 3 groups gives `epsilon` = $1.5/2$, and the denominator agrees: $130.5/174$ is also .75. The ratio of the two reported degrees of freedom, $130.5/1.5 = 87$, is the number of subjects the analysis used less the number of groups (for `between_within`, that difference divided by the between-subjects degrees of freedom, `df_numerator/df_num_within`); when it points to a different number of subjects than the study reports, that number is the N to supply. An implied `epsilon` above 1 or below its lower bound (see Arguments) means the reported degrees of freedom are not a Greenhouse-Geisser or Huynh-Feldt correction of this effect; a Huynh-Feldt estimate above 1 is reported as the uncorrected degrees of freedom, which is `epsilon` = 1. `ss_buc_mixed_anova_general` then applies the approximation of Muller and Barton (1989) to both studies: the prior F is treated as noncentral F with both degrees of freedom and the noncentrality parameter multiplied by `epsilon`, the publication threshold is the critical value at the corrected degrees of freedom, and the planned study's power is computed the same way. It applies to the effects with a within-subjects component (`within_only` and `between_within`); a `between_only` effect is tested on subject means and does not involve sphericity. For `between_within` it is the ϵ of the within-subjects component's contrasts. The returned `ncp_adjusted` stays on the sphericity scale, so it is comparable across values of `epsilon`, and the planned test's corrected degrees of freedom are the `df_effect` and `df_error` attributes. One `epsilon` describes both studies, because it is a property of the covariance matrix of the repeated measures, which the planned study shares

with the prior one; the reported value is the prior study's estimate of it, and re-running across the values you find plausible shows how much the plan depends on it. The value used is echoed as the epsilon row of the result, 1 when sphericity is assumed.

The correction under epsilon is an approximation. Below sphericity the distribution of a within-subjects F depends on the whole covariance matrix of the repeated measures, and on how the effect lines up with it, not on ϵ alone. In simulations of raw data (one-way designs of 4 and 6 levels with ϵ from .45 to .86, prior studies of 12 to 40 subjects published when their Greenhouse-Geisser test reached $p < .05$, and a 2 x 4 split-plot interaction), planning from each published study with its own reported Greenhouse-Geisser estimate at an assurance of .80 gave a plan reaching .80 power for the Greenhouse-Geisser test in .70 to .80 of replications when the effect followed the dominant pattern of variation (a linear trend over autocorrelated occasions), and in .92 to .99 when it lay in the least variable contrasts. Ignoring the departure (epsilon = 1) reached it in .26 to .64 of replications for the trend-like effects and in .70 to .87 for the effects in the least variable contrasts. The shortfall does not shrink as the prior study grows, and is not larger near the lower bound of ϵ . The planned study's power is approximated the same way. Separate simulations of planned one-way studies of 10 to 30 subjects, with 4, 6, or 8 levels, measured the real power of the Greenhouse-Geisser test, whose estimate of ϵ is biased downward. For a trend-like effect (a linear trend over occasions with first-order autoregressive correlations of .3 to .9, ϵ from .50 to .94) it fell below actual_power by about .04 to .08 with four levels and .06 to .13 with six, the larger shortfalls for the smaller plans; a planned study analyzed with the Huynh-Feldt correction tracks actual_power more closely. Even when sphericity holds and epsilon = 1, a planned study analyzed with the Greenhouse-Geisser correction anyway, as some software does by default, has less power than actual_power, the power of the uncorrected test: about .02 to .06 less with four levels, .03 to .11 with six, and .04 to .15 with eight, for plans of 30 down to 10 subjects (the Huynh-Feldt test is within about .02). When a paper reports both corrections, supply the Greenhouse-Geisser one: the Huynh-Feldt estimate is larger, and planning from it attained about .02 to .09 less, the larger differences from the smaller prior studies. When sphericity in fact holds but the prior study reported a Greenhouse-Geisser correction anyway, supplying it is conservative, with a median plan 7 to 34 percent larger.

The planning effect size and other power software. Here f_{planning} is Cohen's f relative to the effect's own error term, whose variance already reflects the correlation among the repeated measures, so $f^2/(1 + f^2)$ is the population value of the partial eta squared a paper reports for the effect. It is not the f that power functions for repeated measures designs usually take, such as `ss_power_split_plot_anova()` in the sibling package DMAR: those standardize the effect by the standard deviation of the raw scores and take the correlation among the measures as a separate argument, so giving them f_{planning} , with any correlation, plans a different study. For a design with one between-subjects factor and one within-subjects factor of b levels, the value that reproduces this plan in `ss_power_split_plot_anova()`, with the correlation set to 0 and the same epsilon, is f_{planning} times $\sqrt{k/b}$, where k is the error degrees of freedom per subject (1 for between_only, $b - 1$ for the other two effects). The between_only effect is included: its f_{planning} is relative to the variance of subject means, not of raw scores.

Value

An object of class `bucss_plan`: 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 power of the plan returned, that is, of the returned

sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic, at the prior study's sample size (for a prior N that does not divide evenly, at the balanced sample size the planner substitutes for it), and on the sphericity scale for every epsilon (see Details); and the planning effect size, `f_planning`, Cohen's f for the effect tested, relative to its own error term, $\sqrt{\lambda/(kN)}$, where λ is `ncp_adjusted`, N the total of the balanced prior design it belongs to, and k the error degrees of freedom per subject, 1 for `between_only`, `df_numerator` for `within_only`, and `df_num_within` for `between_within`, so $f^2/(1 + f^2)$ is the population partial eta squared for the effect; like `ncp_adjusted` it is on the sphericity scale, and for epsilon below 1 the power statement that follows holds under the same Muller and Barton (1989) approximation, with the corrected degrees of freedom and noncentrality parameter; it is not the f that power functions for repeated measures take (see Details).

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through assurance). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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. `print.bucss_plan` shows the design, the effect, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112. doi:10.1007/BF02289823

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, 1, 69–82. doi:10.3102/10769986001001069

Muller, K. E., & Barton, C. N. (1989). Approximate power for repeated-measures ANOVA lacking sphericity. *Journal of the American Statistical Association*, 84, 549–555. doi:10.1080/01621459.1989.10478802

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

# A within-subjects effect with 2 numerator degrees of freedom, from a
# prior study of 90 subjects in 3 groups reported with Greenhouse-Geisser
# corrected degrees of freedom of 1.5 and 130.5: the correction is 1.5/2
# (under sphericity, the plan is 41 per group).
ss_buc_mixed_anova_general(F_observed = 8, N = 90, df_numerator = 2,
  num_groups = 3, effect = "within_only", epsilon = .75)

# 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, together with the planning effect size, the effect size the plan is built on, as a standardized mean.

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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an alpha_prior of 1.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses desired_power, when each is planned from a prior study published under the rule alpha_prior describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so assurance = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With alpha_prior = 1 and assurance = .80 the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed alpha_planned: a test's power is never below its

Type I error rate, so a value at or below `alpha_planned` is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless `alpha_planned` is below .01.

Details

The one-sample t test and the dependent (paired) t test are the same test applied to a single column of scores (Maxwell, Delaney, & Kelley, 2027): 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_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `npc_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the t statistic (the noncentrality parameter of the prior study's noncentral t); and the planning effect size, `std_mean_planning`, the standardized mean (the distance of the mean from its null value in standard deviations, a magnitude), $\lambda/\sqrt{N}$, where λ is `npc_adjusted` and N the prior sample size; it is the `dz_planning` of `ss_buc_paired_t` under the one-sample name.

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through assurance). It is `npc_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `npc_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the

test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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](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. doi:[10.1080/00273171.2017.1289361](https://doi.org/10.1080/00273171.2017.1289361)
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:[10.1037/1082989X.8.3.305](https://doi.org/10.1037/1082989X.8.3.305)
- 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. doi:[10.1080/03610929608831787](https://doi.org/10.1080/03610929608831787)

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, together with the planning effect size, the effect size the plan is built on, as Cohen’s *f*.

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 <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 (groups) for the single between-subjects factor.
alpha_prior	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with <i>p</i> below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every <i>p</i> value is at most 1, so any result would have been published whatever its <i>p</i> value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an alpha_prior of 1.

alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses desired_power, when each is planned from a prior study published under the rule alpha_prior describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so assurance = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With alpha_prior = 1 and assurance = .80 the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed alpha_planned: a test's power is never below its Type I error rate, so a value at or below alpha_planned is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless alpha_planned is below .01.

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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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, $\text{assurance} > .5$ can be selected, which corresponds to selecting the noncentrality parameter associated with the $(1 - \text{assurance})$ quantile of the likelihood distribution.

alpha_prior is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below alpha_prior reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting $\text{alpha_prior} = .20$ assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set $\text{alpha_prior} = 1$, which may also be written $\text{alpha_prior} = \text{"No Publication Filter"}$; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by assurance , remains.

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 $\text{alpha_prior} = 1$ 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.

$\text{ss_buc_one_way_anova}$ 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. The planning effect size belongs to the same balanced design as the reported noncentrality parameter, so an ordinary power analysis at it can return a sample size a little smaller than the one reported here (never a larger one), because the reported size is the larger of the two rounded designs' requirements. For a prior study with unequal group sizes, or a total sample size that does not divide evenly, the largest supportable assurance is computed at the balanced design the planner plans from, so it can be below $1 - p/\alpha$ at the study's own p value. 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_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic, at the prior study's sample size (for a prior N that does not divide evenly, at the balanced sample size the planner substitutes for it; see Details); and the planning effect size, `f_planning`, Cohen's f , $\sqrt{\lambda/N}$, where λ is `ncp_adjusted` and N the total of the balanced prior design it belongs to (see Details), so $\lambda = Nf^2$, the convention of power analysis for between-subjects designs, and $f^2/(1 + f^2)$ is the population partial eta squared for the effect (for a one-way design, eta squared).

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through assurance). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787

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, together with the planning effect size, the effect size the plan is built on, as a standardized mean of the paired differences (Cohen's d_z).

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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an alpha_prior of 1.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses desired_power, when each is planned from a prior study published under the rule alpha_prior describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so assurance = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With alpha_prior = 1 and assurance = .80 the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed alpha_planned: a test's power is never below its Type I error rate, so a value at or below alpha_planned is not allowable, and

the function stops rather than return a plan. 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 (.01), which is then refused unless `alpha_planned` is below .01.

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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

`alpha_prior` is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below `alpha_prior` reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting `alpha_prior = .20` assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set `alpha_prior = 1`, which may also be written `alpha_prior = "No Publication Filter"`; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by `assurance`, remains.

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 of the difference scores (Cohen's d_z , defined under Value) 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_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the t statistic (the noncentrality parameter of the prior study's noncentral t); and the planning effect size, `dz_planning`, Cohen's (1988) d_z (a magnitude): the mean of the paired differences divided by the standard deviation of those differences, the subscript z naming the difference score (Cohen writes $Z = X - Y$); here $\lambda/\sqrt{N}$, where λ is `ncp_adjusted` and N the prior number of pairs.

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through assurance). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression

planners the predictor frame, the unit, and the ceiling beneath the table; tidy() and glance() give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

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 the Squared Multiple Correlation 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, together with the planning effect size, the effect size the plan is built on, as a squared multiple correlation.

Usage

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

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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an alpha_prior of 1.

alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses <code>desired_power</code> , when each is planned from a prior study published under the rule <code>alpha_prior</code> describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so <code>assurance = .5</code> plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below <code>.5</code> is not recommended. The term follows Kelley and Maxwell (2003), who used <code>assurance</code> in sample size planning for the probability that a confidence interval will be no wider than desired. With <code>alpha_prior = 1</code> and <code>assurance = .80</code> the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as <code>.80</code>); 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 <code>.80</code>); the same applies to <code>desired_power</code> .
desired_power	Desired level of statistical power for the planned study: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed <code>alpha_planned</code> : a test's power is never below its Type I error rate, so a value at or below <code>alpha_planned</code> is not allowable, and the function stops rather than return a plan. 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 (<code>.01</code>), which is then refused unless <code>alpha_planned</code> is below <code>.01</code> .
random_predictors	TRUE if the predictors are sampled along with the outcome, as in most observational research, and FALSE if their values are fixed by design, as in an experiment that assigns them. The default is FALSE, the frame of BUCSS 1.x; see Details. With random predictors (TRUE) the planner uses the exact distribution for sampled predictors that the correlation planner also uses: with one predictor ($p = 1$) it returns exactly what the correlation planner, ss_buc_correlation , returns. The choice is echoed in the result as a <code>random_predictors</code> row valued 1 (TRUE) or 0 (FALSE) and named in the printed footer.

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.

Fixed or random predictors. Whether the predictors' values are fixed by design, as in an experiment in which the researcher sets them, or sampled along with the outcome, as in most observational research, changes neither the estimates nor the distribution of the test statistic when every coefficient

is zero, but it does change the statistic's distribution, and so the power, when they are not all zero (Sampson, 1974; Gatsonis & Sampson, 1989; Kelley & Maxwell, 2003). With `random_predictors = FALSE`, the default, the prior study's statistic and the planned study's power come from the noncentral F distribution, which is exact when the predictor values are fixed; this is the frame of BUCSS 1.x and of Anderson (2021). With `random_predictors = TRUE` the noncentrality parameter is itself random: it is proportional to the predictors' sampled spread, a chi square variable on $N - 1$ degrees of freedom, so the statistic follows a chi square mixture of noncentral F distributions, which is more dispersed. This is the exact distribution of the squared multiple correlation when the predictors are sampled (the general sampling distribution of the multiple correlation coefficient; Fisher, 1928), the one for which Gatsonis and Sampson (1989) give exact power and sample size calculations. The mixture has an exact closed form (the Poisson weights of the noncentral F become negative binomial weights), used for both the correction and the planned study's power. It is exact when the predictors are multivariate normal. In observational research in the behavioral sciences the predictors are typically random (Kelley & Maxwell, 2003), and `random_predictors = TRUE` is then the frame to plan in.

Treating sampled predictors as fixed overstates the planned study's power, by about $1/n$ for a plan of n participants with one predictor, $1.2/n$ with three, and $1.4/n$ with eight to ten (.010 to .014 at $n = 100$, and .05 to .06 at $n = 20$), so the assurance the plans deliver falls a little short of the requested assurance. The delivered assurance is what assurance promises: the proportion of plans, each made from its own published prior study, whose power at the true effect reaches `desired_power`, counting a refusal to plan as meeting the target. In simulations with three multivariate normal predictors, an assurance of .80, and prior studies with power .3 to .8 at the true effect, the fixed frame delivered about .76 to .80 from prior studies of 30 to 40, the lower values from better powered prior studies, and .79 to .80 from prior studies of 100 to 300; `random_predictors = TRUE` delivered .79 to .82 in the same configurations. The random-predictor plan is usually a few participants larger; close to the assurance ceiling, where plans run to thousands, it can be a few percent smaller. The largest assurance a prior result can support is the same in both frames, because at a zero effect both are the central F distribution. With one predictor and `random_predictors = TRUE` this function returns exactly what `ss_buc_correlation` returns for the same prior t (the square root of F_{observed}) and N (the same sample size, power, noncentrality parameter, and ceiling, with the planning effect size stated as R^2 , the square of the correlation), for any prior N of at least 4 (the correlation planner's minimum).

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 (with `random_predictors = TRUE`, of the mixture described above). 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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

`alpha_prior` is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below `alpha_prior` reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting `alpha_prior` = .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set `alpha_prior` = 1, which may also be written `alpha_prior` = "No Publication Filter"; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by assurance, remains.

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 search for the planned sample size starts at $p + 3$, the smallest total that leaves the planned study two error degrees of freedom (as in BUCSS 1.x), so no plan is smaller than that, even when the prior effect is so large that a smaller study would already reach the desired power. A plan at this floor signals a very large prior effect, not a recommended design.

Value

An object of class `bucss_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic at the prior study's sample size; and the planning effect size, `R2_planning`, the population squared multiple correlation, $\lambda/(\lambda + N)$, where λ is `ncp_adjusted` and N the prior sample size; the planned study's power is computed from it at the planned sample size, in the frame `random_predictors` names (with sampled predictors the power statement that follows is in that frame).

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through assurance). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787
- 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
- Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 121, 654–673. doi:10.1098/rspa.1928.0224
- Gatsonis, C., & Sampson, A. R. (1989). Multiple correlation: Exact power and sample size calculations. *Psychological Bulletin*, 106, 516–524. doi:10.1037/00332909.106.3.516
- Sampson, A. R. (1974). A tale of two regressions. *Journal of the American Statistical Association*, 69, 682–689. doi:10.1080/01621459.1974.10480189

Examples

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

# Predictors sampled along with the outcome, as in most observational
# research:
ss_buc_R2(F_observed = 5, N = 150, p = 4, random_predictors = TRUE)

# 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, together with the planning effect size, the effect size the plan is built on, as Cohen's f^2 .

Usage

```
ss_buc_reg_coef(
  t_observed,
  N,
  p,
  alpha_prior = 0.05,
  alpha_planned = 0.05,
  assurance = 0.8,
  desired_power = 0.8,
  random_predictors = FALSE
)
```

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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an alpha_prior of 1.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses desired_power, when each is planned from a prior study published under the rule alpha_prior describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so assurance = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With alpha_prior = 1 and assurance = .80 the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed alpha_planned: a test's power is never below its Type I error rate, so a value at or below alpha_planned is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless alpha_planned is below .01.
random_predictors	TRUE if the predictors are sampled along with the outcome, as in most observational research, and FALSE if their values are fixed by design, as in an experiment that assigns them. The default is FALSE, the frame of BUCSS 1.x; see Details. With random predictors (TRUE) the planner uses the exact distribution for sampled predictors that the correlation planner also uses: with one predictor it returns exactly what the correlation planner, ss_buc_correlation , returns. The choice is echoed in the result as a random_predictors row valued 1 (TRUE) or 0 (FALSE) and named in the printed footer.

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.

Fixed or random predictors. Whether the predictors' values are fixed by design, as in an experiment in which the researcher sets them, or sampled along with the outcome, as in most observational research, changes neither the estimates nor the distribution of the test statistic when the coefficient is zero, but it does change the statistic's distribution, and so the power, when the coefficient is not zero (Sampson, 1974; Gatsonis & Sampson, 1989; Kelley & Maxwell, 2003). With `random_predictors = FALSE`, the default, the prior study's statistic and the planned study's power come from the noncentral F distribution, which is exact when the predictor values are fixed; this is the frame of BUCSS 1.x and of Anderson (2021). With `random_predictors = TRUE` the noncentrality parameter is itself random: it is proportional to the sampled spread of the tested predictor once the other predictors are accounted for, a chi square variable on $N - p$ degrees of freedom, so the statistic follows a chi square mixture of noncentral F distributions, which is more dispersed. (When every predictor is tested, the mixture is the general sampling distribution of the multiple correlation coefficient that Fisher, 1928, derived; for one coefficient it is that distribution for the tested predictor once the other predictors are accounted for.) The mixture has an exact closed form (the Poisson weights of the noncentral F become negative binomial weights), used for both the correction and the planned study's power. It is exact when the tested predictor is normally distributed around its linear regression on the other predictors, which may themselves be fixed by design (dummy codes for assigned groups, say). In observational research in the behavioral sciences the predictors are typically random (Kelley & Maxwell, 2003), and `random_predictors = TRUE` is then the frame to plan in.

Treating sampled predictors as fixed overstates the planned study's power, by about $1/n$ for a plan of n participants with one predictor, $1.9/n$ with three, and $4/n$ with eight (.010, .019, and .040 at $n = 100$, and $4.8/n$ with eight at $n = 20$), so the assurance the plans deliver can fall short of the requested assurance. The delivered assurance is what assurance promises: the proportion of plans, each made from its own published prior study, whose power at the true effect reaches `desired_power`, counting a refusal to plan as meeting the target. In simulations with multivariate normal predictors, an assurance of .80, and prior studies of 20 to 100 participants (25 to 100 with eight predictors) with power .3 to .8 at the true effect, the fixed frame delivered about .75 to .80 with one or three predictors and .75 to .82 with eight, the lowest values from small prior studies with power near .8; `random_predictors = TRUE` delivered .79 to .82 in the same configurations.

Which frame plans the larger study depends on the other predictors and on how far the plan outgrows the prior study, because the fixed frame scales the prior study's noncentrality parameter by N while the tested predictor's sampled spread has $N - p$ degrees of freedom. With one predictor the random-predictor plan is usually a few participants larger (and at most a few percent smaller). With more predictors the two plans are within a few participants, the random-predictor plan usually the larger, when the plan is no more than about twice the prior N ; beyond that the random-predictor plan is usually smaller, and the more the plan outgrows the prior study, the closer the ratio of the two plans comes to $(N - p)/N$, so the random-predictor plan is smaller by at most about p/N : about 15 percent with three predictors and 40 percent with eight from a prior study of 20, and 3 and 8

percent from a prior study of 100 (83 rather than 98 participants for $t = 2.5$ from a prior study of 30 with eight predictors and $\alpha_{\text{prior}} = 1$, and 623 rather than 624 in the examples below). The largest assurance a prior result can support is the same in both frames, because at a zero effect both are the central F distribution. With one predictor and $\text{random_predictors} = \text{TRUE}$ this function returns exactly what `ss_buc_correlation` returns for the same prior t and N (the same sample size, power, noncentrality parameter, and ceiling, with the planning effect size stated as f^2 rather than as the correlation), for any prior N of at least 4 (the correlation planner's minimum).

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 (with $\text{random_predictors} = \text{TRUE}$, of the mixture described above). The denominator is the power of the test, which serves to truncate the distribution. In the single predictor case with fixed predictors, this formula reduces to the density of a truncated noncentral t distribution. (See Taylor & Muller, 1996, Equation 2.1; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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 $\text{assurance} = .5$ (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.

α_{prior} is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below α_{prior} reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting $\alpha_{\text{prior}} = .20$ assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set $\alpha_{\text{prior}} = 1$, which may also be written $\alpha_{\text{prior}} = \text{"No Publication Filter"}$; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by assurance, remains.

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 α_{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 α_{prior} at a value greater than .05, which raises the ceiling. Note that α_{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 α_{prior} is the only remaining lever. It is possible to correct for uncertainty only by setting $\alpha_{\text{prior}} = 1$ 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 search for the planned sample size starts at $p + 3$, the smallest total that leaves the planned study two error degrees of freedom (as in BUCSS 1.x), so no plan is smaller than that, even when the prior effect is so large that a smaller study would already reach the desired power. A plan at this floor signals a very large prior effect, not a recommended design.

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_observed = -3` plans exactly the sample size that `t_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_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic (the square of the prior t) at the prior study's sample size; and the planning effect size, `f2_planning`, Cohen's f^2 for the coefficient, λ/N , where λ is `ncp_adjusted` and N the prior sample size; the planned study's power is computed from it at the planned sample size, in the frame `random_predictors` names (with sampled predictors it is the population f^2 of that frame, and the power statement that follows is in that frame).

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through assurance). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787
- 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
- Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 121, 654–673. doi:10.1098/rspa.1928.0224
- Gatsonis, C., & Sampson, A. R. (1989). Multiple correlation: Exact power and sample size calculations. *Psychological Bulletin*, 106, 516–524. doi:10.1037/00332909.106.3.516
- Sampson, A. R. (1974). A tale of two regressions. *Journal of the American Statistical Association*, 69, 682–689. doi:10.1080/01621459.1974.10480189

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

# Predictors sampled along with the outcome, as in most observational
# research:
ss_buc_reg_coef(t_observed = 3, N = 150, p = 3, random_predictors = TRUE)

# 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	<i>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</i>
------------------	--

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, together with the planning effect size, the effect size the plan is built on, as Cohen's f^2 .

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,
  random_predictors = FALSE
)
```

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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1

	when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an <code>alpha_prior</code> of 1.
<code>alpha_planned</code>	Alpha level α assumed for the planned study.
<code>assurance</code>	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses <code>desired_power</code> , when each is planned from a prior study published under the rule <code>alpha_prior</code> describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so <code>assurance</code> = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With <code>alpha_prior</code> = 1 and <code>assurance</code> = .80 the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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 <code>desired_power</code> .
<code>desired_power</code>	Desired level of statistical power for the planned study: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed <code>alpha_planned</code> : a test's power is never below its Type I error rate, so a value at or below <code>alpha_planned</code> is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless <code>alpha_planned</code> is below .01.
<code>random_predictors</code>	TRUE if the predictors are sampled along with the outcome, as in most observational research, and FALSE if their values are fixed by design, as in an experiment that assigns them. The default is FALSE, the frame of BUCSS 1.x; see Details. With random predictors (TRUE) the planner uses the exact distribution for sampled predictors that the correlation planner also uses: with one predictor, tested (<code>p</code> = 1, <code>p_joint</code> = 1), it returns exactly what the correlation planner, ss_buc_correlation , returns. The choice is echoed in the result as a <code>random_predictors</code> row valued 1 (TRUE) or 0 (FALSE) and named in the printed footer.

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.

Fixed or random predictors. Whether the predictors' values are fixed by design, as in an experiment in which the researcher sets them, or sampled along with the outcome, as in most observational research, changes neither the estimates nor the distribution of the test statistic when the tested coefficients are zero, but it does change the statistic's distribution, and so the power, when they are not all zero (Sampson, 1974; Gatsonis & Sampson, 1989; Kelley & Maxwell, 2003). With `random_predictors = FALSE`, the default, the prior study's statistic and the planned study's power come from the noncentral F distribution, which is exact when the predictor values are fixed; this is the frame of BUCSS 1.x and of Anderson (2021). With `random_predictors = TRUE` the noncentrality parameter is itself random: it is proportional to the sampled spread of the tested predictors once the other predictors are accounted for, a chi square variable on $N - 1 - (p - p_{\text{joint}})$ degrees of freedom, so the statistic follows a chi square mixture of noncentral F distributions, which is more dispersed. (When every predictor is tested, the mixture is the general sampling distribution of the multiple correlation coefficient that Fisher, 1928, derived; for a subset it is that distribution for the tested predictors once the other predictors are accounted for.) The mixture has an exact closed form (the Poisson weights of the noncentral F become negative binomial weights), used for both the correction and the planned study's power. It is exact when the tested predictors are multivariate normal around their linear regression on the other predictors, which may themselves be fixed by design (dummy codes for assigned groups, say). In observational research in the behavioral sciences the predictors are typically random (Kelley & Maxwell, 2003), and `random_predictors = TRUE` is then the frame to plan in.

Treating sampled predictors as fixed overstates the planned study's power, by about $1/n$ to $1.4/n$ for a plan of n participants when every predictor is tested, and more the more predictors are left untested: $2/n$ for two of four and $3.8/n$ for two of eight (at $n = 100$), so the assurance the plans deliver can fall short of the requested assurance. The delivered assurance is what assurance promises: the proportion of plans, each made from its own published prior study, whose power at the true effect reaches `desired_power`, counting a refusal to plan as meeting the target. In simulations with multivariate normal predictors, an assurance of .80, and prior studies of 40 to 100 participants with power .3 to .8 at the true effect, testing two of four or five predictors or three of six, the fixed frame delivered about .76 to .81, the lower values from smaller, better powered prior studies; `random_predictors = TRUE` delivered .79 to .81 in the same configurations.

Which frame plans the larger study depends on the untested predictors and on how far the plan outgrows the prior study, because the fixed frame scales the prior study's noncentrality parameter by N while the tested predictors' sampled spread has $N - 1 - (p - p_{\text{joint}})$ degrees of freedom. With every predictor tested the random-predictor plan is usually a few participants larger (and at most a few percent smaller). With predictors left untested the two plans are within a few participants, the random-predictor plan usually the larger, when the plan is no more than about twice the prior N ; beyond that the random-predictor plan is usually smaller, and the more the plan outgrows the prior study, the closer the ratio of the two plans comes to $(N - 1 - (p - p_{\text{joint}}))/N$, so the random-predictor plan is smaller by at most about $(1 + p - p_{\text{joint}})/N$: 518 rather than 600 participants for $F = 6$ from a prior study of 30 testing two of six predictors, where the limit is about 17 percent, and, with two of eight predictors tested, a limit of 35 percent from a prior study of 20 and 7 percent from 100. The largest assurance a prior result can support is the same in both frames, because at a zero effect both are the central F distribution. With `p_joint = p` this function returns what `ss_buc_R2` returns, and with `p_joint = 1` what `ss_buc_reg_coef` returns for the square root of F_{observed} , in either frame. So with one predictor, tested ($p = 1$, $p_{\text{joint}} = 1$), and `random_predictors = TRUE`, it returns exactly what `ss_buc_correlation` returns for the same prior t (the square root of F_{observed}) and N (the same sample size, power, noncentrality parameter, and ceiling, with the planning effect size stated as f^2 rather than as the correlation), for any prior N of at least 4 (the

correlation planner's minimum).

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 (with `random_predictors = TRUE`, of the mixture described above). 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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

`alpha_prior` is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below `alpha_prior` reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting `alpha_prior = .20` assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set `alpha_prior = 1`, which may also be written `alpha_prior = "No Publication Filter"`; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by `assurance`, remains.

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 search for the planned sample size starts at $p + 3$, the smallest total that leaves the planned

study two error degrees of freedom (as in BUCSS 1.x), so no plan is smaller than that, even when the prior effect is so large that a smaller study would already reach the desired power. A plan at this floor signals a very large prior effect, not a recommended design.

Value

An object of class `bucss_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see *Details*); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic at the prior study's sample size; and the planning effect size, `f2_planning`, Cohen's f^2 for the set of predictors tested, λ/N , where λ is `ncp_adjusted` and N the prior sample size; the planned study's power is computed from it at the planned sample size, in the frame `random_predictors` names (with sampled predictors it is the population f^2 of that frame, and the power statement that follows is in that frame).

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through `assurance`). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an `assurance` of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361

Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305

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. doi:10.1080/03610929608831787

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

Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 121, 654–673. doi:10.1098/rspa.1928.0224

Gatsonis, C., & Sampson, A. R. (1989). Multiple correlation: Exact power and sample size calculations. *Psychological Bulletin*, 106, 516–524. doi:10.1037/00332909.106.3.516

Sampson, A. R. (1974). A tale of two regressions. *Journal of the American Statistical Association*, 69, 682–689. doi:10.1080/01621459.1974.10480189

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

# Predictors sampled along with the outcome, as in most observational
# research:
ss_buc_reg_joint(F_observed = 5, N = 150, p = 4, p_joint = 2,
  random_predictors = TRUE)

# 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	Necessary Sample Size to Reach Desired Power for a One- or Two-Way Within-Subjects ANOVA Using a Publication Bias and Uncertainty Correction Procedure
-----------------	--

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, together with the planning effect size, the effect size the plan is built on, as Cohen's f .

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,
  epsilon = 1
)
```

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>levels_A</code>	Number of levels for Factor A.
<code>levels_B</code>	Number of levels for Factor B, which is NULL for a single-factor design.
<code>effect</code>	Effect of most interest to the planned study: main effect of A (<code>factor_A</code>), main effect of B (<code>factor_B</code>), or interaction (<code>interaction</code>). Case-insensitive shorthand is also accepted but not recommended in scripts: "A" (or "a") for <code>factor_A</code> , "B" for <code>factor_B</code> , and "AxB" (or "AB") for interaction.
<code>alpha_prior</code>	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below <code>alpha_prior</code> . Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by <code>assurance</code>) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an <code>alpha_prior</code> of 1.
<code>alpha_planned</code>	Alpha level α assumed for the planned study.
<code>assurance</code>	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses <code>desired_power</code> , when each is planned from a prior study published under the rule <code>alpha_prior</code> describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the

	prior statistic, so assurance = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With <code>alpha_prior = 1</code> and <code>assurance = .80</code> the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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 <code>desired_power</code>.
<code>desired_power</code>	Desired level of statistical power for the planned study: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed <code>alpha_planned</code>: a test's power is never below its Type I error rate, so a value at or below <code>alpha_planned</code> is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless <code>alpha_planned</code> is below .01.
<code>epsilon</code>	The sphericity parameter ϵ of the effect tested, from 1 divided by the effect's numerator degrees of freedom (its lower bound) to 1 (the default, which assumes sphericity). For a prior study reported with Greenhouse-Geisser or Huynh-Feldt corrected degrees of freedom, it is the reported numerator degrees of freedom divided by the uncorrected ones. An effect with one numerator degree of freedom (a two-level factor) is always spherical and takes only 1. The same value describes the planned study, and it is echoed as the <code>epsilon</code> row of the result. See Details.

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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

`alpha_prior` is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below `alpha_prior` reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting `alpha_prior = .20` assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set `alpha_prior = 1`, which may also be written `alpha_prior = "No Publication Filter"`; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by `assurance`, remains.

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.

Sphericity. The default `epsilon = 1` assumes sphericity, as the uncorrected within-subjects test does. When a prior study reports Greenhouse-Geisser (1959) or Huynh-Feldt (1976) corrected degrees of freedom, supply `epsilon`, the reported numerator degrees of freedom divided by the uncorrected ones. For example, $F(1.73, 50.2)$ for a three-level within-subjects factor gives `epsilon = 1.73/2`, about .865, and the denominator agrees: $50.2/58$ is also about .865. The ratio of the two reported degrees of freedom, $50.2/1.73 = 29.0$, is one less than the number of subjects the analysis used, here the study's 30; when it points to a different number (after dropping subjects with missing data, say), that number is the N to supply. An implied `epsilon` above 1 or below its lower bound (see Arguments) means the reported degrees of freedom are not a Greenhouse-Geisser or Huynh-Feldt correction of this effect; a Huynh-Feldt estimate above 1 is reported as the uncorrected degrees of freedom, which is `epsilon = 1`. `ss_buc_rm_anova` then applies the approximation of Muller and Barton (1989) to both studies: the prior F is treated as noncentral F with both degrees of freedom and the noncentrality parameter multiplied by `epsilon`, the publication threshold is the critical value at the corrected degrees of freedom, and the planned study's power is computed the same

way. Each within-subjects effect has its own ϵ , so supply the one reported for the effect you plan from. The returned `ncp_adjusted` stays on the sphericity scale, so it is comparable across values of epsilon, and the planned test's corrected degrees of freedom are the `df_effect` and `df_error` attributes. One epsilon describes both studies, because it is a property of the covariance matrix of the repeated measures, which the planned study shares with the prior one; the reported value is the prior study's estimate of it, and re-running across the values you find plausible shows how much the plan depends on it. The value used is echoed as the epsilon row of the result, 1 when sphericity is assumed.

The correction under epsilon is an approximation. Below sphericity the distribution of a within-subjects F depends on the whole covariance matrix of the repeated measures, and on how the effect lines up with it, not on ϵ alone. In simulations of raw data (one-way designs of 4 and 6 levels with ϵ from .45 to .86, prior studies of 12 to 40 subjects published when their Greenhouse-Geisser test reached $p < .05$, and a 2 x 4 split-plot interaction), planning from each published study with its own reported Greenhouse-Geisser estimate at an assurance of .80 gave a plan reaching .80 power for the Greenhouse-Geisser test in .70 to .80 of replications when the effect followed the dominant pattern of variation (a linear trend over autocorrelated occasions), and in .92 to .99 when it lay in the least variable contrasts. Ignoring the departure (epsilon = 1) reached it in .26 to .64 of replications for the trend-like effects and in .70 to .87 for the effects in the least variable contrasts. The shortfall does not shrink as the prior study grows, and is not larger near the lower bound of ϵ . The planned study's power is approximated the same way. Separate simulations of planned one-way studies of 10 to 30 subjects, with 4, 6, or 8 levels, measured the real power of the Greenhouse-Geisser test, whose estimate of ϵ is biased downward. For a trend-like effect (a linear trend over occasions with first-order autoregressive correlations of .3 to .9, ϵ from .50 to .94) it fell below `actual_power` by about .04 to .08 with four levels and .06 to .13 with six, the larger shortfalls for the smaller plans; a planned study analyzed with the Huynh-Feldt correction tracks `actual_power` more closely. Even when sphericity holds and epsilon = 1, a planned study analyzed with the Greenhouse-Geisser correction anyway, as some software does by default, has less power than `actual_power`, the power of the uncorrected test: about .02 to .06 less with four levels, .03 to .11 with six, and .04 to .15 with eight, for plans of 30 down to 10 subjects (the Huynh-Feldt test is within about .02). When a paper reports both corrections, supply the Greenhouse-Geisser one: the Huynh-Feldt estimate is larger, and planning from it attained about .02 to .09 less, the larger differences from the smaller prior studies. When sphericity in fact holds but the prior study reported a Greenhouse-Geisser correction anyway, supplying it is conservative, with a median plan 7 to 34 percent larger.

The planning effect size and other power software. Here `f_planning` is Cohen's f relative to the effect's own error term, whose variance already reflects the correlation among the repeated measures, so $f^2/(1 + f^2)$ is the population value of the partial eta squared a paper reports for the effect. It is not the f that power functions for repeated measures designs usually take, such as `ss_power_rm_anova()` in the sibling package DMAR: those standardize the effect by the standard deviation of the raw scores and take the correlation among the measures as a separate argument, so giving them `f_planning`, with any correlation, plans a different study. For a one-way design with a levels, the value that reproduces this plan in `ss_power_rm_anova()`, with the correlation set to 0 and the same epsilon, is `f_planning` times $\sqrt{(a - 1)/a}$.

Value

An object of class `bucss_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); ncp_adjusted, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic, at the prior study's sample size, and on the sphericity scale for every epsilon (see Details); and the planning effect size, f_planning, Cohen's f for the effect tested, relative to its own error term, $\sqrt{\lambda/(kN)}$, where λ is ncp_adjusted, N the prior number of subjects, and k the effect's numerator degrees of freedom (the error degrees of freedom per subject), so $f^2/(1 + f^2)$ is the population partial eta squared for the effect; like ncp_adjusted it is on the sphericity scale, and for epsilon below 1 the power statement that follows holds under the same Muller and Barton (1989) approximation, with the corrected degrees of freedom and noncentrality parameter; it is not the f that power functions for repeated measures take (see Details).

A planning effect size (the row whose name ends in _planning) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through alpha_prior, and the uncertainty allowed for through assurance). It is ncp_adjusted restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates ncp_adjusted, actual_power is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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. print.bucss_plan shows the design, the effect, the unit, and the ceiling beneath the table; tidy() and glance() give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112. doi:10.1007/BF02289823
- Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, 1, 69–82. doi:10.3102/10769986001001069
- Muller, K. E., & Barton, C. N. (1989). Approximate power for repeated-measures ANOVA lacking sphericity. *Journal of the American Statistical Association*, 84, 549–555. doi:10.1080/01621459.1989.10478802

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

# A prior study reported with Greenhouse-Geisser corrected degrees of
# freedom, 1.73 and 50.2, for a three-level factor with 30 subjects: the
# correction is 1.73/2, and the plan is larger than under sphericity (92
# subjects).
ss_buc_rm_anova(F_observed = 6.5, N = 30, levels_A = 3, epsilon = 1.73 / 2)

# 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, together with the planning effect size, the effect size the plan is built on, as Cohen's f .

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,
  epsilon = 1
)
```

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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below <code>alpha_prior</code> . Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by <code>assurance</code>) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an <code>alpha_prior</code> of 1.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses <code>desired_power</code> , when each is planned from a prior study published under the rule <code>alpha_prior</code> describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so <code>assurance = .5</code> plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below <code>.5</code> is not recommended. The term follows Kelley and Maxwell (2003), who used <code>assurance</code> in sample size planning for the probability that a confidence interval will be no wider than desired. With <code>alpha_prior = 1</code> and <code>assurance = .80</code> the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). Enter it as a proportion in (0, 1) or as a percentage greater than 1 (for example, 80 is read as <code>.80</code>); 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 <code>.80</code>); the same applies to <code>desired_power</code> .

desired_power	Desired level of statistical power for the planned study: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed alpha_planned: a test's power is never below its Type I error rate, so a value at or below alpha_planned is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless alpha_planned is below .01.
epsilon	The sphericity parameter ϵ of the effect tested, from 1/df_numerator (its lower bound) to 1 (the default, which assumes sphericity). For a prior study reported with Greenhouse-Geisser or Huynh-Feldt corrected degrees of freedom, it is the reported numerator degrees of freedom divided by the uncorrected ones. An effect with df_numerator = 1 is always spherical and takes only 1. The same value describes the planned study, and it is echoed as the epsilon row of the result. See Details.

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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

alpha_prior is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below alpha_prior reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting alpha_prior = .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set alpha_prior = 1, which may also be written alpha_prior = "No Publication Filter"; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not

to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by assurance, remains.

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.

Sphericity. The default `epsilon` = 1 assumes sphericity, as the uncorrected within-subjects test does. When a prior study reports Greenhouse-Geisser (1959) or Huynh-Feldt (1976) corrected degrees of freedom, supply `epsilon`, the reported numerator degrees of freedom divided by the uncorrected ones. For example, $F(1.73, 50.2)$ for an effect with 2 numerator degrees of freedom gives `epsilon` = $1.73/2$, about .865, and the denominator agrees: $50.2/58$ is also about .865. The ratio of the two reported degrees of freedom, $50.2/1.73 = 29.0$, is one less than the number of subjects the analysis used, here the study's 30; when it points to a different number (after dropping subjects with missing data, say), that number is the N to supply. An implied `epsilon` above 1 or below its lower bound (see Arguments) means the reported degrees of freedom are not a Greenhouse-Geisser or Huynh-Feldt correction of this effect; a Huynh-Feldt estimate above 1 is reported as the uncorrected degrees of freedom, which is `epsilon` = 1. `ss_buc_rm_anova_general` then applies the approximation of Muller and Barton (1989) to both studies: the prior F is treated as noncentral F with both degrees of freedom and the noncentrality parameter multiplied by `epsilon`, the publication threshold is the critical value at the corrected degrees of freedom, and the planned study's power is computed the same way. Each within-subjects effect has its own ϵ , so supply the one reported for the effect you plan from. The returned `ncp_adjusted` stays on the sphericity scale, so it is comparable across values of `epsilon`, and the planned test's corrected degrees of freedom are the `df_effect` and `df_error` attributes. One `epsilon` describes both studies, because it is a property of the covariance matrix of the repeated measures, which the planned study shares with the prior one; the reported value is the prior study's estimate of it, and re-running across the values you find plausible shows how much the plan depends on it. The value used is echoed as the `epsilon` row of the result, 1 when sphericity is assumed.

The correction under `epsilon` is an approximation. Below sphericity the distribution of a within-subjects F depends on the whole covariance matrix of the repeated measures, and on how the effect lines up with it, not on ϵ alone. In simulations of raw data (one-way designs of 4 and 6 levels with

ϵ from .45 to .86, prior studies of 12 to 40 subjects published when their Greenhouse-Geisser test reached $p < .05$, and a 2 x 4 split-plot interaction), planning from each published study with its own reported Greenhouse-Geisser estimate at an assurance of .80 gave a plan reaching .80 power for the Greenhouse-Geisser test in .70 to .80 of replications when the effect followed the dominant pattern of variation (a linear trend over autocorrelated occasions), and in .92 to .99 when it lay in the least variable contrasts. Ignoring the departure ($\epsilon = 1$) reached it in .26 to .64 of replications for the trend-like effects and in .70 to .87 for the effects in the least variable contrasts. The shortfall does not shrink as the prior study grows, and is not larger near the lower bound of ϵ . The planned study's power is approximated the same way. Separate simulations of planned one-way studies of 10 to 30 subjects, with 4, 6, or 8 levels, measured the real power of the Greenhouse-Geisser test, whose estimate of ϵ is biased downward. For a trend-like effect (a linear trend over occasions with first-order autoregressive correlations of .3 to .9, ϵ from .50 to .94) it fell below `actual_power` by about .04 to .08 with four levels and .06 to .13 with six, the larger shortfalls for the smaller plans; a planned study analyzed with the Huynh-Feldt correction tracks `actual_power` more closely. Even when sphericity holds and $\epsilon = 1$, a planned study analyzed with the Greenhouse-Geisser correction anyway, as some software does by default, has less power than `actual_power`, the power of the uncorrected test: about .02 to .06 less with four levels, .03 to .11 with six, and .04 to .15 with eight, for plans of 30 down to 10 subjects (the Huynh-Feldt test is within about .02). When a paper reports both corrections, supply the Greenhouse-Geisser one: the Huynh-Feldt estimate is larger, and planning from it attained about .02 to .09 less, the larger differences from the smaller prior studies. When sphericity in fact holds but the prior study reported a Greenhouse-Geisser correction anyway, supplying it is conservative, with a median plan 7 to 34 percent larger.

The planning effect size and other power software. Here `f_planning` is Cohen's f relative to the effect's own error term, whose variance already reflects the correlation among the repeated measures, so $f^2/(1 + f^2)$ is the population value of the partial eta squared a paper reports for the effect. It is not the f that power functions for repeated measures designs usually take, such as `ss_power_rm_anova()` in the sibling package DMAR: those standardize the effect by the standard deviation of the raw scores and take the correlation among the measures as a separate argument, so giving them `f_planning`, with any correlation, plans a different study. For a one-way design with a levels (`df_numerator = a - 1`), the value that reproduces this plan in `ss_power_rm_anova()`, with the correlation set to 0 and the same epsilon, is `f_planning` times $\sqrt{(a - 1)/a}$.

Value

An object of class `bucss_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the F statistic, at the prior study's sample size, and on the sphericity scale for every epsilon (see Details); and the planning effect size, `f_planning`, Cohen's f for the effect tested, relative to its own error term, $\sqrt{\lambda/(kN)}$, where λ is `ncp_adjusted`, N the prior number of subjects, and k is `df_numerator` (the error degrees of freedom per subject), so $f^2/(1 + f^2)$ is the population partial eta squared for the effect; like `ncp_adjusted` it is on the sphericity scale, and for epsilon below 1 the power statement that follows holds under the same Muller and Barton (1989) approximation, with the corrected degrees of freedom and noncentrality parameter; it is not the f that power functions for

repeated measures take (see Details).

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through assurance). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an assurance of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112. doi:10.1007/BF02289823
- Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, 1, 69–82. doi:10.3102/10769986001001069
- Muller, K. E., & Barton, C. N. (1989). Approximate power for repeated-measures ANOVA lacking sphericity. *Journal of the American Statistical Association*, 84, 549–555. doi:10.1080/01621459.1989.10478802

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

# A prior study reported with Greenhouse-Geisser corrected degrees of
# freedom, 1.4 and 40.6, for an effect with 2 numerator degrees of freedom
# and 30 subjects: the correction is 1.4/2 (under sphericity, the plan is
# 46 subjects).
ss_buc_rm_anova_general(F_observed = 8, N = 30, df_numerator = 2,
  epsilon = .7)

# 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. At an assurance of .80 the default of 1000 gives attainment_with_refusals a Monte Carlo standard error near .013, and 10000 about .004; both are reported. The standard error of attainment, which counts only the issued plans, is larger when many prior studies are refused.
seed	Optional random number seed. NULL (the default) draws from the caller's own random number stream, advancing it as any simulation does. When a seed is supplied, the caller's random number state and generator are restored on exit (and no state is left behind if the caller had none). A seed reproduces a result only under the same random number generators (see RNGkind).

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, the probability that a study's test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). 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 planner's stop when the corrected noncentrality parameter is zero, or, for a design the planner reduces to a balanced one, its stop when a published statistic is not significant at the balanced design's degrees of freedom) 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. For a prior study that the planner reduces to a balanced design it can land above, because the planner is then conservative: about .84 at an assurance of .80 for prior groups of 10 and 90 in the independent *t* test.

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 $(\text{assurance} - \text{refusal_rate}) / (1 - \text{refusal_rate})$, or more when `attainment_with_refusals` lands above assurance.

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. For the regression planners the simulated prior studies follow the plan's own `random_predictors` setting, so audit a plan in the frame your research actually uses. For a within-subjects or split-plot plan made with `epsilon` below 1, the prior studies are drawn under the same approximation to a departure from sphericity that the planner corrects under, so the audit checks the method's internal consistency, not the approximation itself; the planner's help page reports how well the approximation holds against simulated raw data.

The package-wide verification ships with the package. This function audits one plan; the harness that verifies the planners themselves, including checks against raw simulated data that do not share the method's distributional assumptions (bivariate normal samples for the correlation planner, multivariate normal samples for a regression coefficient with `random_predictors = TRUE`, two-group

normal samples for the Welch planner), is installed with BUCSS and can be rerun in a few seconds with `source(system.file("tools", "verify_by_simulation.R", package = "BUCSS"))`.

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 each with its Monte Carlo standard error, the sample size the true effect requires, and the quartiles of the sample sizes the simulated plans called for. The design, the effect tested (for designs with one), the sample size unit, the requested assurance, and the requested desired_power travel as attributes. If every simulated prior study is refused, attainment, its standard error, and the quartiles are NA, with a warning.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787

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)
ncp <- plan$value[plan$term == "ncp_adjusted"]
ncp

# Take the corrected estimate at face value as the truth. (200 replications
```

```
# keep the example quick: the rate counting refusals is good to about .03,
# the rate over issued plans to about .06. A reported audit deserves 10000.)
ss_buc_sensitivity(plan, true_ncp = ncp, replications = 200, seed = 113)

# And ask what happens if the truth is smaller than the correction assumed.
ss_buc_sensitivity(plan, true_ncp = .6 * ncp, replications = 200, seed = 113)
```

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, together with the planning effect size, the effect size the plan is built on, as a standardized mean difference.

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	The significance level α a prior result had to reach to be published (more generally, to reach you): the correction assumes you see only results with p below alpha_prior. Enter a number above 0 and at most 1. A value of 1, which may also be written "No Publication Filter", means no publication filter: every p value is at most 1, so any result would have been published whatever its p value, there is no publication bias to correct, and only the correction for uncertainty (set by assurance) remains, as for a pilot study. Anderson, Kelley, and Maxwell (2017, p. 1552) note that this level, their α_P , can be set equal to 1 when publication bias is thought not to apply. The string is read without regard to case, and the result echoes it as an alpha_prior of 1.
alpha_planned	Alpha level α assumed for the planned study.
assurance	Desired level of assurance: the long run proportion of planned studies whose statistical power at the true effect reaches or surpasses desired_power, when each is planned from a prior study published under the rule alpha_prior describes. The adjusted noncentrality parameter is the lower limit of a one-sided confidence interval at this level, computed from the publication-truncated distribution of the prior statistic, so assurance = .5 plans from a median unbiased estimate (correcting for publication bias only), a larger value adds the correction for uncertainty, and a value below .5 is not recommended. The term follows Kelley and Maxwell (2003), who used assurance in sample size planning for the probability that a confidence interval will be no wider than desired. With alpha_prior = 1 and assurance = .80 the plan corresponds to safeguard power, which plans for the lower limit of a two-sided 60 percent confidence interval around the prior effect size rather than for the effect size itself (see BUCSS-package for how, and for where the two differ). 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: the probability that its test rejects the null hypothesis when the effect is present (Maxwell, Delaney, & Kelley, 2027). It must exceed alpha_planned: a test's power is never below its Type I error rate, so a value at or below alpha_planned is not allowable, and the function stops rather than return a plan. 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 (.01), which is then refused unless alpha_planned is below .01.

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 (Welch, 1947) is not exactly distributed as a noncentral t : its denominator degrees of freedom are estimated from the data by the Welch-Satterthwaite formula (Satterthwaite, 1946; Welch, 1947), 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, and computes the planned study's power the same way. **A minimum planned sample size of 10 per group is imposed, because of estimation accuracy issues at smaller sample sizes.** With fewer per group the approximation is not accurate enough: the real test estimates its two variances, and so its degrees of freedom, from too few observations for the noncentral t to describe it, and the approximation overstates its power, badly so for the smallest (a plan of 2 per group that it puts at .82 has real power near .50). From 10 per group, at an `alpha_planned` of .05, the approximation is accurate: across ratios from 1 to 5, the power of the real Welch test is within .006 of what this function reports for every plan it returns, and within .003 from about 12 per group at a target power of .80, and its Type I error stays between .047 and .054 against a nominal .05 across ratios from 1 to 3. At a stricter `alpha_planned` the approximation overstates the power of the smallest plans by more, at 10 per group by up to about .012 at .01, .016 at .005, and .025 at .001, and it is within .006 again from 14, 16, and 20 per group, respectively. When the target power is reached with fewer than 10 per group, the planner returns 10, whose power exceeds the target.

The correction of the prior study's statistic uses the same approximation and holds up well unless a prior group is very small. Simulating prior studies as raw data, published by the real Welch test, the corrected parameter fell at or below the truth in about .79 to .81 of them at an assurance of .80 when each prior group had 10 to 20 participants (a refusal counted as at or below). With 5 to 8 in the smaller prior group it ranged from about .78 to .86 when the true ratio was supplied, and fell as low as about .73 when the ratio was taken from the prior study's own standard deviations. One consequence of reading the prior study on the Welch-Satterthwaite degrees of freedom the supplied ratio implies, rather than on those its data produced: when the prior groups are small and unequal in size or in spread, including at the default ratio of 1, a prior result its paper reported as significant can fall short of the critical value here and be refused as nonsignificant. Such a result is typically close to the publication threshold, where it could support little assurance in any case.

The assumption that matters is the ratio itself. One `sd_ratio` describes both studies: the prior study, whose statistic is read under it, and the planned study. Getting it wrong costs far more than the distributional approximation does. For the example below, the plan built assuming equal variances (213 per group) delivers about .76 power rather than .80 if the ratio is in fact 2 in both studies, and about .43 if the prior study's groups were equally variable but the planned study's ratio is 2, with the mean difference held fixed in units of the first group's standard deviation. Re-running across a range of plausible ratios, and planning for the least favorable one you find credible, guards against the first kind of error. The second, a planned study more unequal than the prior one, is not something the prior result can speak to, and needs its own allowance. The required sample size does not always rise as the two standard deviations move apart: when the group assumed to be more variable is the smaller one it often falls, though not always (for the prior study of the example below it rises).

When the ratio is 1 and the prior groups were the same size, this planner returns what `ss_buc_independent_t` returns, except for the minimum of 10 per group imposed here. 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 usually larger when one prior group is more than about twice the size of the other;

when the two sizes are close it is more often the same, and it can be smaller, because the equal-variance planner rounds the harmonic mean down and up to whole group sizes and returns the larger of the two plans.

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; Anderson, Kelley, & Maxwell, 2017, Equation 3; 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.

alpha_prior is the significance level a prior result had to reach to be published: the correction assumes that only results with p values below alpha_prior reach the literature, and so reach you. Different α levels represent differing amounts of publication bias. For example, setting alpha_prior = .20 assumes that studies with p values less than .20 are published, whereas those with larger p values are not, which reflects less severe publication bias than the default of .05. If the previous study of interest has not been subjected to publication bias (e.g., a pilot study), set alpha_prior = 1, which may also be written alpha_prior = "No Publication Filter"; Anderson, Kelley, and Maxwell (2017, p. 1552) note that it can be set equal to 1 when publication bias is thought not to apply. Because no p value exceeds 1, this means no publication filter at all: any result would have been published whatever its p value, so there is no publication bias to correct, and only the correction for uncertainty, set by assurance, remains.

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 10 percent fewer participants in total at a ratio of 2, and about 20 percent fewer at a ratio of 3.

Value

An object of class `bucss_plan`: 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 power of the plan returned, that is, of the returned sample size at the returned noncentrality parameter (see Details); `ncp_adjusted`, the publication bias and uncertainty adjusted prior study noncentrality parameter, on the scale of the t statistic (the noncentrality parameter of the prior study's noncentral t); and the planning effect size, `smd_planning`, a standardized mean difference (a magnitude) whose standardizer is the root mean square of the two groups' standard deviations (Bonett, 2008), $\lambda\sqrt{(1/n_1 + r^2/n_2)/((1 + r^2)/2)}$, where λ is `ncp_adjusted`, n_1 and n_2 are the prior group sizes, and r is the ratio of standard deviations; with n per group in the planned study its noncentrality parameter is this value times $\sqrt{n/2}$, as for the equal-variance test, and with a ratio of 1 it is Cohen's δ .

A planning effect size (the row whose name ends in `_planning`) is an adjusted effect size: the value, in the design's usual effect size metric, used to plan the sample size given the situation at hand (the prior study's result, the publication bias assumed through `alpha_prior`, and the uncertainty allowed for through `assurance`). It is `ncp_adjusted` restated in that metric and is what the plan is built on, not an estimate of the population effect to report: above an `assurance` of .5 it is deliberately conservative (a lower confidence limit), and at .5 it is the median unbiased estimate. Because it restates `ncp_adjusted`, `actual_power` is the ordinary power of a study of the returned sample size when the population effect is the planning effect size.

Rows echoing the planning inputs follow the design results, 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` where the test has a denominator) travel on attributes. `print.bucss_plan` shows the design, for the regression planners the predictor frame, the unit, and the ceiling beneath the table; `tidy()` and `glance()` give the compact and wide views, the latter with every attribute as a column.

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. doi:10.1080/00273171.2017.1289361
- Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8, 305–321. doi:10.1037/1082989X.8.3.305
- 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. doi:10.1080/03610929608831787
- Bonett, D. G. (2008). Confidence intervals for standardized linear contrasts of means. *Psychological Methods*, 13, 99–109. doi:10.1037/1082989X.13.2.99
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2, 110–114. doi:10.2307/3002019
- Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, 34, 28–35. doi:10.1093/biomet/34.12.28

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), [3](#)
BUCSS-package, [3](#)
bucss_tidiers, [5](#)

glance.bucss_plan (bucss_tidiers), [5](#)

planning_sentence, [7](#)
plot, [3](#)
plot.bucss_plan, [8](#)

RNGkind, [98](#)

ss_buc_ancova, [3](#), [10](#)
ss_buc_chisq_diff, [3](#), [15](#)
ss_buc_correlation, [3](#), [20](#), [71](#), [72](#), [76](#), [78](#),
[82](#), [83](#)
ss_buc_factorial_anova, [3](#), [25](#), [30](#), [61](#)
ss_buc_factorial_anova_general, [3](#), [12](#),
[14](#), [26](#), [30](#), [61](#)
ss_buc_independent_t, [3](#), [35](#), [43](#), [100](#), [103](#),
[106](#)
ss_buc_manova, [3](#), [40](#)
ss_buc_mixed_anova, [3](#), [44](#)
ss_buc_mixed_anova_general, [3](#), [51](#)
ss_buc_one_sample_t, [3](#), [57](#)
ss_buc_one_way_anova, [3](#), [12](#), [26](#), [30](#), [61](#)
ss_buc_paired_t, [3](#), [59](#), [60](#), [65](#)
ss_buc_R2, [3](#), [22](#), [70](#), [83](#)
ss_buc_reg_coef, [3](#), [22](#), [25](#), [75](#), [83](#)
ss_buc_reg_joint, [3](#), [22](#), [81](#)
ss_buc_rm_anova, [3](#), [86](#)
ss_buc_rm_anova_general, [3](#), [92](#)
ss_buc_sensitivity, [3](#), [4](#), [10](#), [98](#)
ss_buc_smd (ss_buc_independent_t), [35](#)
ss_buc_smd_paired (ss_buc_paired_t), [65](#)
ss_buc_welch_t, [3](#), [99](#), [101](#)

tidy.bucss_plan (bucss_tidiers), [5](#)

uniroot, [3](#)