Power Analysis and Sample Size Determination

H. Hoijtink[1], Qianrao Fu[1]

[1]Department of Methodology and Statistics
Utrecht University H.Hoijtink@uu.nl

2018

NHST

It all begins with the null-hypothesis, for example,

$$H_0 : \mu_1 = \mu_2 \tag{1}$$

which is evaluated against the alternative hypothesis

$$H_a : \mu_1, \mu_2 \tag{2}$$

using data x_i for $i = 1, ..., N_1$ persons in Group 1 and $i = 1, ..., N_2$ persons in Group 2.

Note that, it is assumed that in each group the data have a normal distribution with a group dependent mean, and a variance σ^2.

The p-value

p-value

The p-value is the probability of the sample difference in means (or larger) under the assumption that $H_0 : \mu_1 = \mu_2$ is true.

$\overline{x}_1$	$\overline{x}_2$	p-value
0	0	1
0	.2	.65
0	.5	.22
0	.8	.03

According to Sir R.A. Fisher, the p-value is a measure of evidence against H_0, that is, the smaller the p-value, the larger the evidence against H_0.

The Neyman-Pearson Approach to Evaluating p-values

In the sample $\rightarrow$	$p < \alpha$	$p > \alpha$
In the population H_0 is true	Type I error	
In the population H_a is true	power	Type II error

A cut-off value is introduced (α) to which the p-value is compared in order to decide whether "H_0 cannot be rejected" ($p > \alpha$) or whether "H_0 can be rejected" ($p < \alpha$).

Before elaborating this procedure, terminology has to be introduced.

Terminology

Cohen's *d*

Cohen's d = $\frac{\overline{x}_1 - \overline{x}_2}{\hat{sd}}$, that is , the number of standard deviations two sample means differ from each other. It is an effect size measure, that is, a standardized measure of the difference between two means.

Note that, sd^2 is the estimate of σ^2.

Note furthermore, that according to Cohen, $d = .2, .5, .8$, denotes small, medium, and large effect sizes, respectively.

More Terminology

Type I error and α

A Type I error occurs if H_0 is incorrectly rejected. The probability of a Type I error is denoted by α. The value commonly used for α is .05.

Type II error and β

A Type II error occurs if H_0 is incorrectly not rejected. The probability of a Type I error is denoted by β. The value commonly used for β is .20.

Still More Terminology

Power

Power is the probability of correctly rejecting H_0. It is equal to $1 - \beta$, that is, a value commonly used is .80.

The larger the sample size per group, the larger the power, the table on the next slide shows how a power analysis can be executed, that is, what should the sample size be in order to obtain a power of .80.

Power Analysis

Sample size per group required to obtain a power of .80.

	Effect Size Cohen's d		
α	small = .20	medium = .50	large =.80
.01	586	95	38
.05	393	64	26
.10	310	50	20

- effect sizes are often between .2 and .5
- after choosing α and *d* the sample size per group can be found in the table
- when using Bonferroni kind of corrections, the α level has to be adjusted (this will be discussed later in this lecture)

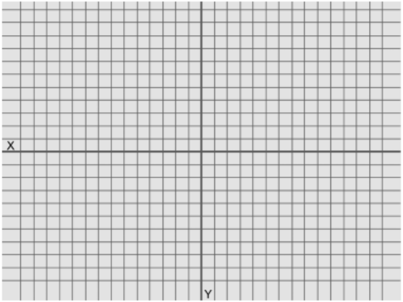
1. Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

One Way ANOVA

Can spatial distance influence the judgments of individuals? Students are "primed" with spatial "closeness" or "distance" with the help of a Cartesian distance plot. Subsequently, they are asked to quantify the strength of their bond with brothers, sisters, parents, and hometown on a 1-7 Likert scale.

1. Williams, L.E. and Bargh, J.A. (2008). Keeping One's Distance. The Influence of Spatial Distance Cues on Affect and Evaluation. *Psychological Sciene, 19*, 302-308. http://dx.doi.org/10.1111/j.1467-9280.2008.02084.x

One Way ANOVA



The coordinate system using for priming. The students had to localize the following coordinates:
(2, 4) and (3,1) in the “close” condition,
(8, 3) and (6,5) in the “intermediate” condition,
(12, 10) and (11, 8) in the “distant” condition.

One Way ANOVA

Students were given a questionnaire on which the had to rate the strength of their bond to brothers, sisters, parents, and home-town on a 1 (weak) to 7 (strong) Likert scale. By averaging these score an index for emotional attachment was obtained.

The core hypotheses for this one-way ANOVA are

$$H_0 : \mu_{close} = \mu_{intermediate} = \mu_{distant} \tag{3}$$

versus

$$H_a : \text{ not } H_0 \tag{4}$$

With respect to the data: 28 students were randomly assigned to each of the three conditions.

One Way ANOVA

Results as presented in Williams and Bargh (2008):

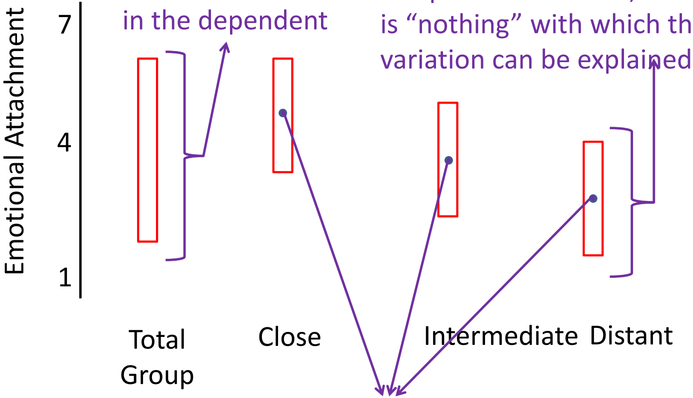
1. H_0, p = .01 , $\eta^2 = .11$
2. $\overline{m}_{distant} = 4.86$, $\overline{m}_{intermediate} = 5.27$ en $\overline{m}_{close} = 5.61$

One Way ANOVA

Explaining the Results

Total variation in the dependent

Uneplained variation, there is “nothing” with which this variation can be explained.



Explained variance (by group). The larger the differences between the means, the larger the explained variance.

η^2 is the proportion of the total variance explained by group:

$$\eta^2 = \frac{\text{Explained Variance}}{\text{Total Variance}}$$

η^2 is an effect size measure, the larger η^2 the larger the differences between the group means.

The p-value is equal to .01, that is, smaller than .05, conequently, H_0 is rejected. But what if the p-value would have been larger than .05?

One Way ANOVA

Explaining the Results

If the p-value is larger than .05, there are two options:

1. In the population $\mu_{close} = \mu_{intermediate} = \mu_{distant}$, that is, H_0 should not be rejected.
2. There are differences between the group means, but there was not enough power to reject H_0.

If, before the data are collected, a power analysis is executed, it can be determined which of these two options is the most valid.

Sample sizes per group necessary to obtain a power of .80

	Effect Size: η^2		
α	small = .01	medium = .06	large =.16
.01	464	76	30
.05	322	52	21
.10	258	41	17

One Way ANOVA

Post-Hoc Tests

Knowing that H_0 has to be rejected is not the end of the story.
The remaining question is which pairs of means are different.
This question can be answered by executing post-hoc tests.

Post-hoc tests compare each pair of means. With three groups three post-hoc tests are executed:

$$H_{01}: \mu_{close} = \mu_{intermediate} \text{ versus } H_a: \mu_{close} \neq \mu_{intermediate} \quad (5)$$

$$H_{02}: \mu_{intermediate} = \mu_{distant} \text{ versus } H_a: \mu_{intermediate} \neq \mu_{distant} \quad (6)$$

$$H_{03}: \mu_{close} = \mu_{distant} \text{ versus } H_a: \mu_{close} \neq \mu_{distant} \quad (7)$$

However, if three tests are executed, there are three opportunities to "commit" a Type I error. The probability of "one or more" Type I errors will no longer be .05, but larger.

Below you find examples of number of tests, and the corresponding probability of 1 or more (α level) and 2 or more Type I errors.

number of tests	probability of 1 or more incorrect rejections of H_0	probability of 2 or more incorrect rejections of H_0
1	.05	
2	.0975	.0025
3	.1426	.0073
...		
10	.4012	.0861

One Way ANOVA

Post-Hoc Tests

The Bonferronie approach can be used (more efficient approaches exist) to control the probability of one or more incorrect rejections of H_0, that is, to keep is at maximally .05.

With three tests Bonferroni works as follows:

Compare each p-value not to .05 but to .05/3.

Note that, if we divide .05 by 3, we have to redo the power analysis! Lets do that!

One Way ANOVA

Post-Hoc Tests

comparison	p-value	η^2 or Cohen's d
all three groups	.01	.11
close - intermediate	.12	.41
close - distant	.003	.83
intermediate -distant	.12	.41

Problems with Power Analysis

- How to choose the effect size?
- How to choose the α level, that is, .05 or .05 divided by the number of hypotheses tested? How to choose the number of hypotheses tested?
- If, when analyzing the real data set, p > α, "it is all over" one cannot update :-((

Sample Size Determination

Sample size determination to achieve what?

- To obtain a certain degree of support ($BF_{0a} = 4.5$) if the null is true and to obtain a certain degree of support ($BF_{a0} = 4.5$) when the alternative is true
- To obtain small Type I and Type II errors !! if the Bayes factor is used as a decision criterion: $BF_{0a} > 1$ decide in favor of H_0, $BF_{0a} < 1$ decide in favor of H_a
- Like the previous option but then with a region of "indecision"

1. Fu, Q., Hoijtink, H., and Moerbeek, M. (unpublished). Sample size determination for the Bayesian t-test. Retrievable from the `bain` website.

Sample Size Determination Procedure

- Choose $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1, \mu_2$ or $H_a : \mu_1 > \mu_2$
- Choose Bayesian t-test or Bayesian Welch's test
- Choose an effect size d for H_a
- Choose the required degree of support, e.g., BF_{0a} is at least 4.5 if the null is true and BF_{a0} is at least 4.5 when the alternative is true (how to do this?)
- Determine the N per group such that:
    - When many data sets with N persons per group are simulated from a population in which H_0 is true, the median BF_{0a} across all data sets is equal to 4.5.
    - When many data sets with N persons per group are simulated from a population in which H_a with effect size d is true, the median BF_{a0} across all data sets is equal to 4.5.

Using the Bayes Factor as a Measure of Support

Table 3

When effect sizes $d = 0.2$ and $d = 0.5$, sample size N, the corresponding median BF_{0i} and 60% intervals of BF_{0i} (the top row) and median BF_{i0} and 60% intervals of BF_{i0} (the bottom row) for Student's t-test ($\sigma^2 = 1$) / Welch's t-test($\sigma_1^2 = 1.33$, $\sigma_2^2 = 0.67$).

d			0.2		0.5	
variances			equal	unequal	equal	unequal
$J = 1$	medBF=5	two-sided	506 25.39 (14.25, 30.86) 5.32 (0.50, 104.00)	505 25.40 (14.12, 30.81) 5.02 (0.49, 104.80)	65 9.05 (4.92, 11.02) 5.34 (0.64, 91.43)	65 9.03 (4.92, 11.06) 5.27 (0.65, 93.02)
		one-sided	429 28.98 (12.38, 50.68) 5.25 (0.63, 95.74)	429 28.57 (12.29, 50.62) 5.40 (0.61, 98.52)	52 10.18 (4.42, 18.08) 5.16 (0.78, 68.29)	52 10.16 (4.43, 18.05) 5.13 (0.78, 68.82)
	medBF=10	two-sided	588 27.34 (15.48, 33.21) 10.80 (0.84, 265.40)	588 27.47 (15.62, 33.19) 10.65 (0.85, 263.30)	80 10 (5.37, 12.21) 12.81 (1.11, 275.70)	80 10.03 (5.41, 12.23) 13.00 (1.13, 279.30)
		one-sided	506 31.37 (14.02, 55.45) 10.64 (0.99, 208.00)	505 31.46 (13.90, 55.66) 10.03 (0.97, 209.50)	65 11.31 (4.88, 20.08) 10.65 (1.26, 182.80)	65 11.34 (4.97, 20.12) 10.52 (1.28, 186)
$J = 2$	medBF=5	two-sided	470 17.29 (9.43, 21.04) 5.41 (0.56, 96.61)	463 17.19 (9.40, 20.82) 5.04 (0.55, 93.79)	59 6.10 (3.30, 7.44) 5.28 (0.75, 85.97)	59 6.10 (3.34, 7.45) 5.29 (0.74, 85.89)
		one-sided	389 19.69 (8.75, 35.03) 5.01 (0.63, 72.08)	390 19.71 (8.46, 34.29) 5.08 (0.68, 81.74)	46 6.86 (2.97, 12.02) 5.08 (0.90, 60.77)	46 6.86 (3.01, 11.96) 5.05 (0.91, 59.16)
	medBF=10	two-sided	546 18.60 (10.39, 22.62) 10.11 (0.91, 232.90)	545 18.62 (10.52, 22.60) 10.10 (0.91, 238.80)	158 10.02 (5.56, 12.18) 1444 (51.19, 100800)	158 10.03 (5.69, 12.18) 1435 (51.83, 102300)
		one-sided	470 21.41 (9.48, 37.92) 10.82 (1.12, 193.20)	463 21.12 (9.24, 37.34) 10.07 (1.09, 187.60)	101 10.10 (4.43, 17.86) 106.20 (7.11, 3457)	101 10.08 (4.40, 17.93) 109.10 (7.25, 3458)
$J = 3$	medBF=5	two-sided	447 13.73 (7.70, 16.70) 5.14 (0.60, 92.65)	438 13.63 (7.60, 16.56) 5.21 (0.58, 83.60)	60 5.07 (2.78, 6.12) 6.98 (0.95, 96.41)	60 5.08 (2.76, 6.12) 6.82 (0.94, 99.91)
		one-sided	362 15.32 (6.76, 26.95) 5.04 (0.71, 68.39)	367 15.62 (6.95, 27.57) 5.02 (0.70, 67.59)	42 5.32 (2.35, 9.37) 5.12 (0.98, 54.67)	42 5.34 (2.32, 9.31) 5.10 (0.99, 54.71)
	medBF=10	two-sided	519 14.77 (8.29, 18.00) 10.05 (0.94, 218.80)	519 14.83 (8.32, 18.01) 10.19 (0.93, 217.70)	237 10.05 (5.62, 12.20) 228900 (3272, 3.59e+07)	236 10.02 (5.51, 12.16) 206300 (3090, 3.39e+07)
		one-sided	447 17.15 (7.47, 30.21) 10.27 (1.17, 185.30)	438 16.61 (7.53, 29.58) 10.41 (1.14, 167.20)	148 10.01 (4.31, 17.45) 2018 (72.18, 128300)	148 10.03 (4.31, 17.44) 2056 (74.20, 126700)

Using the Bayes Factor as a Binary Decision Criterion

In the sample →	$\mathrm{BF}_{0a} < 1$	$\mathrm{BF}_{0a} > 1$
In the population H_0 is true	Type I error	
In the population H_a is true	power	Type II error

The Bayes factor is compared to a cut-off value of 1 in order to decide whether H_0 or H_a is preferred.

Note that, N per group is always determined using the required degree of support in terms of a median Bayes factor. However, associated with this are Type I and Type II errors if the Bayes factor is used to make a decision.

Using the Bayes Factor as a Binary Decision Criterion

Table 5

Type I (p_1) and Type II (p_2) error rates for Student's/Welch's t-test, effect sizes (d) 0.2 and 0.5, median BF values of 5 and 10, two-sided and one-sided testing, and J = 1, 2, 3.

d			0.2				0.5			
variances			equal		unequal		equal		unequal	
p_1 and p_2			p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2
J = 1	medBF=5	two-sided	0.01	0.29	0.01	0.29	0.03	0.26	0.03	0.26
		one-sided	0.01	0.26	0.01	0.26	0.03	0.24	0.04	0.24
	medBF=10	two-sided	0.01	0.22	0.01	0.22	0.02	0.19	0.03	0.19
		one-sided	0.01	0.20	0.01	0.20	0.03	0.17	0.03	0.17
J = 2	medBF=5	two-sided	0.01	0.28	0.02	0.28	0.05	0.25	0.05	0.25
		one-sided	0.02	0.26	0.02	0.26	0.06	0.22	0.06	0.22
	medBF=10	two-sided	0.01	0.21	0.01	0.21	0.03	0.02	0.03	0.02
		one-sided	0.01	0.19	0.02	0.19	0.04	0.04	0.04	0.04
J = 3	medBF=5	two-sided	0.02	0.27	0.02	0.28	0.06	0.21	0.06	0.21
		one-sided	0.02	0.26	0.02	0.25	0.07	0.20	0.07	0.20
	medBF=10	two-sided	0.02	0.21	0.02	0.21	0.02	0.00	0.02	0.00
		one-sided	0.02	0.18	0.02	0.18	0.04	0.01	0.04	0.01

Using the Bayes Factor as a Binary Decision Criterion

Note that, the Type I and Type II errors can be computed *before* the real empirical data have been observed!

After the real empirical data have been observed, we can compute the Posterior Model Probabilities, which lead to the Bayesian errors, that is, the probability of an incorrect decision after the data are observed!
Both errors can serve are role in research ... or ... are the first ... or the second ... more important?

Using the Bayes Factor as a Trichotomous Decision Criterion

In the sample $\rightarrow$	BF < 1/3	1/3 < BF < 3	BF > 3
In pop. H_0 is true	Type I error	weak	
In pop. H_a is true	power	weak	Type II error

The Bayes factor is compared to a cut-off values of 1/3 and 3 in order to decide whether H_0 or H_a is preferred, or that the verdict is "indecision". The probability of a Type I error is denoted by p_1^M, of a Type II error by p_2^M, , and the probabilty of weak evidence by p^w.

Using the Bayes Factor as a Trichotomous Decision

Table 7

Misleading evidence (p_1^M, p_2^M) and weak evidence (p^w) probabilities with respect to cut-off value 3 for Student's/Welch's t-test, effect sizes 0.2 and 0.5, median BF values of 5 and 10, two-sided and one-sided testing, and J = 1, 2, 3.

d			0.2						0.5					
variances			equal			unequal			equal			unequal		
misleading and weak evidence probability			p_1^M	p^w	p_2^M	p_1^M	p^w	p_2^M	p_1^M	p^w	p_2^M	p_1^M	p^w	p_2^M
$J = 1$	medBF=5	two-sided	0.00	0.15	0.15	0.00	0.15	0.16	0.01	0.20	0.11	0.01	0.20	0.11
		one-sided	0.00	0.16	0.13	0.00	0.17	0.13	0.01	0.23	0.08	0.01	0.23	0.08
	medBF=10	two-sided	0.00	0.13	0.11	0.00	0.13	0.11	0.01	0.17	0.08	0.01	0.17	0.08
		one-sided	0.00	0.14	0.09	0.00	0.14	0.09	0.01	0.19	0.05	0.01	0.19	0.05
$J = 2$	medBF=5	two-sided	0.00	0.16	0.14	0.01	0.17	0.14	0.01	0.24	0.09	0.01	0.24	0.09
		one-sided	0.00	0.18	0.12	0.01	0.19	0.11	0.02	0.27	0.06	0.02	0.27	0.06
	medBF=10	two-sided	0.00	0.15	0.10	0.00	0.14	0.10	0.01	0.06	0.00	0.01	0.06	0.00
		one-sided	0.00	0.15	0.08	0.00	0.15	0.08	0.01	0.11	0.01	0.01	0.11	0.01
$J = 3$	medBF=5	two-sided	0.01	0.18	0.13	0.01	0.18	0.13	0.01	0.26	0.07	0.02	0.26	0.07
		one-sided	0.01	0.19	0.10	0.01	0.19	0.11	0.02	0.30	0.05	0.02	0.30	0.05
	medBF=10	two-sided	0.00	0.16	0.09	0.00	0.16	0.09	0.01	0.04	0.00	0.01	0.04	0.00
		one-sided	0.01	0.16	0.07	0.01	0.16	0.07	0.01	0.07	0.00	0.01	0.08	0.00

Problems with Sample Size Determination

- How to choose the effect size?
- There is no issue with respect to multiple testing (see below)
- If when analyzing the data set, the Bayes factor is "only" 3, one can update :-))

Beyond the Bayesian t-test: one ANOVA example

Work in progress, the next table is a preliminary result

Beyond the Bayesian t-test: one ANOVA example

$H_0: \mu_1 = \mu_2 = \mu_3$ $(\mathbf{0}, \mathbf{0}, \mathbf{0})$ vs $H_a: \mu_1 < \mu_2 < \mu_3$ medBF=5

f		0.1 (-0.1224745,0, 0.1224745)			0.25 (-0.3061862,0, 0.3061862)			0.4 (-0.4898979,0, 0.4898979)		
variances		Equal var=c(1,1,1)	Unequal var=c(1/2,1,3/2)	Robust var=c(1/2,1,3/2)	Equal var=c(1,1,1)	Unequal var=c(1/2,1,3/2)	Robust var=c(1/2,1,3/2)	Equal var=c(1,1,1)	Unequal var=c(1/2,1,3/2)	Robust var=c(1/2,1,3/2)
J=1	N	391	352	420	42	36	45	11	10	15
	BF_{0a} (CI)	521.04 (158.09, 1333.95)	470.45 (141.70, 1203.52)	574.89 (171.37, 1417.48)	58.78 (17.18, 142.31)	49.42 (14.80, 122.86)	60.94 (17.58, 156.32)	15.21 (4.37, 37.99)	13.85 (3.63, 35.60)	20.82 (5.32, 57.29)
	BF_{a0} (CI)	5.05 (0.31, 139.80)	5.04 (0.32, 139.70)	5.31 (0.33, 167.57)	5.29 (0.56, 109.57)	5.19 (0.55, 102.96)	6.95 (0.60, 189.54)	5.11 (0.67, 81.12)	6.44 (0.76, 155.64)	14.23 (1.10, 947.27)
	I/II	0.00/0.32	0.00/0.33	0.00/0.32	0.02/0.27	0.02/0.28	0.02/0.26	0.05/0.26	0.07/0.24	0.06/0.19
	I/II/III	0.00/0.21/0.00	0.00/0.21/0.01	0.00/0.20/0.01	0.01/0.14/0.03	0.01/0.15/0.04	0.01/0.14/0.04	0.02/0.11/0.13	0.04/0.10/0.14	0.03/0.08/0.10
J=2	N	340	306	360	32	28	35	10	10	10
	BF_{0a} (CI)	228.31 (68.42, 578.35)	208.44 (61.39, 529.88)	242.00 (72.73, 613.86)	21.97 (6.71, 54.86)	19.38 (5.82, 48.10)	23.85 (6.59, 60.24)	6.96 (2.01, 17.17)	6.92 (1.82, 17.80)	6.73 (1.51, 18.98)
	BF_{a0} (CI)	5.14 (0.37, 125.43)	5.02 (0.36, 124.01)	5.18 (0.36, 134.16)	5.31 (0.63, 78.85)	5.30 (0.69, 84.57)	6.79 (0.72, 151.82)	8.46 (1.20, 134.62)	12.89 (1.52, 311.28)	14.99 (1.24, 922.53)
	I/II	0.00/0.31	0.01/0.31	0.00/0.31	0.04/0.26	0.04/0.25	0.04/0.24	0.12/0.18	0.13/0.15	0.15/0.17
	I/II/III	0.00/0.19/0.01	0.00/0.19/0.01	0.00/0.19/0.01	0.01/0.12/0.08	0.02/0.12/0.10	0.02/0.11/0.09	0.05/0.05/0.23	0.06/0.04/0.23	0.09/0.06/0.22
J=3	N	306	280	335	27	23	30	11	11	11
	BF_{0a} (CI)	136.89 (39.88, 347.95)	128.54 (36.68, 330.03)	153.61 (45.00, 385.65)	12.60 (3.80, 30.70)	10.73 (3.01, 26.85)	13.72 (3.83, 34.85)	5.07 (1.46, 12.66)	5.06 (1.40, 12.91)	5.10 (1.37, 13.18)
	BF_{a0} (CI)	5.04 (0.39, 112.18)	5.06 (0.40, 121.11)	5.34 (0.42, 136.85)	5.19 (0.77, 65.74)	5.18 (0.75, 80.08)	7.09 (0.82, 146.64)	15.32 (2.00, 243.35)	24.63 (2.60, 579.52)	16.86 (1.84, 498.73)
	I/II	0.01/0.31	0.01/0.30	0.01/0.30	0.06/0.24	0.08/0.24	0.07/0.23	0.15/0.11	0.16/0.09	0.16/0.12
	I/II/III	0.00/0.18/0.02	0.00/0.18/0.02	0.00/0.18/0.02	0.02/0.10/0.14	0.03/0.10/0.17	0.03/0.09/0.13	0.05/0.03/0.29	0.07/0.02/0.29	0.07/0.03/0.28

The Role of Sample Size Determination in Bayesian Inference

Situation 1. The population of interest is small, e.g., persons with a rare disease or cognitive disorder. The control and treatment groups will very likely not contain more than 22 persons. Updating is in this situation not an option.

Using sample size determination you find that with 22 persons per group (as will be shown later in the paper in Table 4) you need true a Cohen's d = .8 to obain a Bayes factor of at least 5 if either H_0 or H_1 is true. Since you expect that the effect of the treatment is much smaller than a Cohen's d of .8 you decide not the execute the experiment in this form.

The Role of Sample Size Determination in Bayesian Inference

Situation 1. Situation 2. Next month a survey will start in which 100, currently single, men and women will be tracked for 21 years. Again updating is not an option.

Analogous to *Situation 1 continued* but now you find that you need 65 persons per group to detect a effect size of Cohen's d = .5 with sufficient support. Since you expect that the effect size will be about .5, you proceed with your plans because your sample size is 100 persons per gender group.

The Role of Sample Size Determination in Bayesian Inference

Situation 3. You have to submit your research plans to the (medical) ethical committee. You want to use updating, but both you and they may want an indication of the sample size needed to obtain sufficient support for different effect sizes under the alternative hypothesis. Only with these numbers you can argue that you have sufficient funds and research time to execute you research plan.

The Role of Sample Size Determination in Bayesian Inference

Situation 4. If the alternative hypothesis is true, you expect an effect size Cohen's d = .5. You determine the sample sizes (as will be shown later in Table 3 these are 65 per group) such that BF_{01} is at least 5 when H_0 is true and that BF_{10} is at least 5 when H_1 is true. After collecting data you compute $\mathrm{BF}_{01} = 2.5$. This is a problem because you did not achieve the desired support.

The support you found was $\mathrm{BF}_{01} = 2.5$ in favor of the null-hypothesis. However, you required a support of minimally 5. You can remedy this by updating, that is, increasing your sample size and recomputation of the Bayes factor. The latter is only possible if updating is an option *Sitations 1 and 2* highlight situation where this is not an option.

The Role of Sample Size Determination in Bayesian Inference

Situation 5 Analogous to *Situation 4*, but now you find $\mathrm{BF}_{01} = 11.3$. This is a problem in the sense that you spend more funds and research time than would have been necessary.

You plan and are able to collect the data from 65 persons per group. If your research design permits this (for example, usually yes for experiments and no for panel studies) you can update until you reach the required support (which may be achieved at a sample size smaller than 65 per group) which will save you funds and research time.