Lakens et al (2018) examples using Dienes’ (2008) calculator

Neil McLatchie

19/06/2018

## Dienes’ (2008) Bayes Factor Calculator

Zoltan Dienes’ calculator is located here: <http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/bayes_factor.swf>.

The calculator uses Flash Player which many browsers have recently started blocking. You can either change the settings in your default browser, but the calculator usually works in FireFox.

## Example 1: Martins, Sheppes, Gross and Mather (2016)

To calculate a Bayes factor, we need a model of the data. For Dienes’ (2008) calculator, that means we need to know the raw effect size and its standard error. These are usually quite easy to obtain.

First calculate the mean difference for the effect:

0.338 - 0.321 # young- old

## [1] 0.017

Recall that Martins et al conducted a t-test: t(62) = 0.35, p = .73.

We can use the t-test to calculate the standard error, because Mdif / t = SEM.

0.017 / 0.35 #

## [1] 0.04857143

First we will specify a vague, uninformed model of H1. This model suggests that all outcomes between 0 and 1 are aconsidered fairly plausible, although smaller values are considered more likely than larger values (see Figure 1).

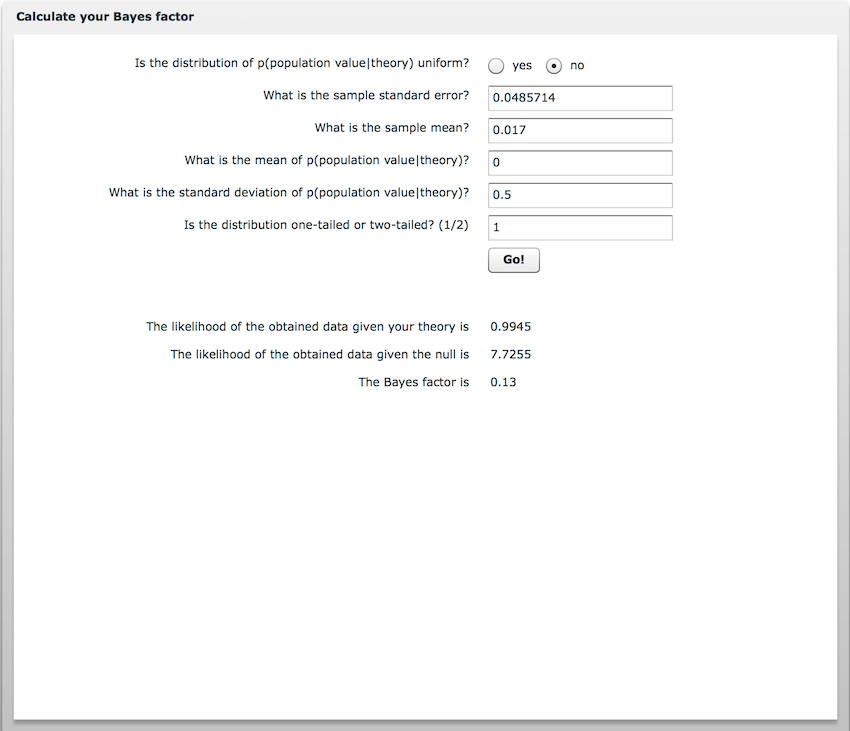


Figure 1. Martins et al, Bayes Factor (Uninformed)

Select “No” at the top, to tell the calcualtor that our prior model is not uniform (i.e., it does not assign equal probability to all possible outcomes).

The top two boxes correspond to our data: enter the obtained SE in to the top one and the Mean difference in to the second box down.

The last three boxes correspond to the model of H1. We specify its mode, standard deviation and whether it is one-tailed or two-tailed. As we believe that smaller effects are more likely than larger effects, we set the mode to 0 (we could set this to a small non-zero number, but it would make very little difference to the result of the Bayes factor, and zero is convenient).

The Standard Deviation of the model of H1 can be thought of as the approximate scale of effect predicted by our theory. Specifying that our model is one-tailed indicates that our hypothesis is directional.

As shown in Figure 1, the obtained Bayes factor provides quite strong evidence for the null hypothesis, . It is also recommended that you specify the robustness regions, which identify the range of predicted effect sizes that would still lead to the same conclusions as our obtained Bayes factor (here, B > 3 counted as evidence for H1, B < 0.33 counted as evidence for H0).

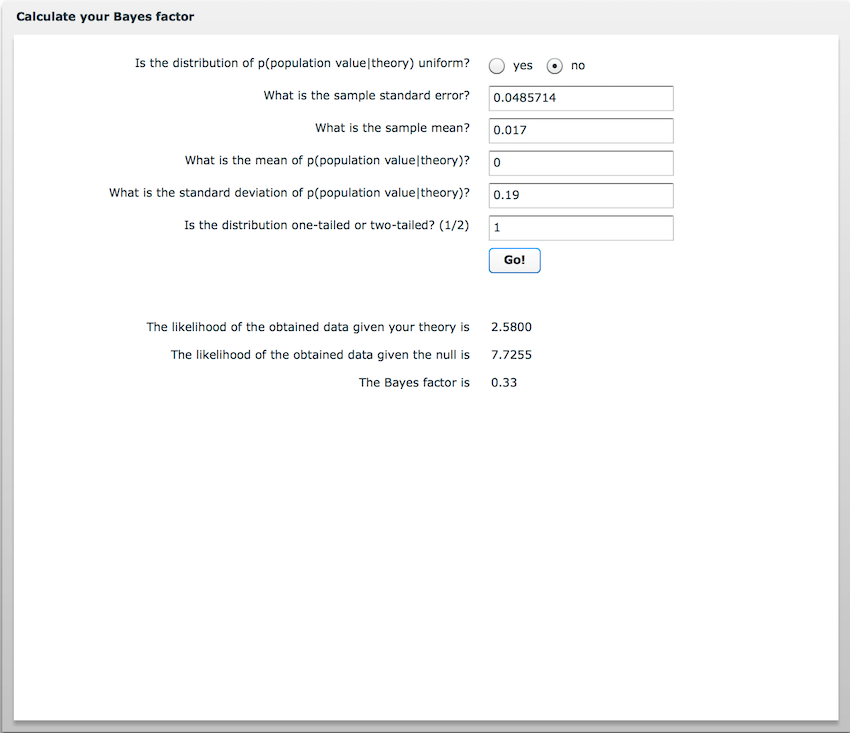


Figure 2. Martins et al, Minimum Robustness Region

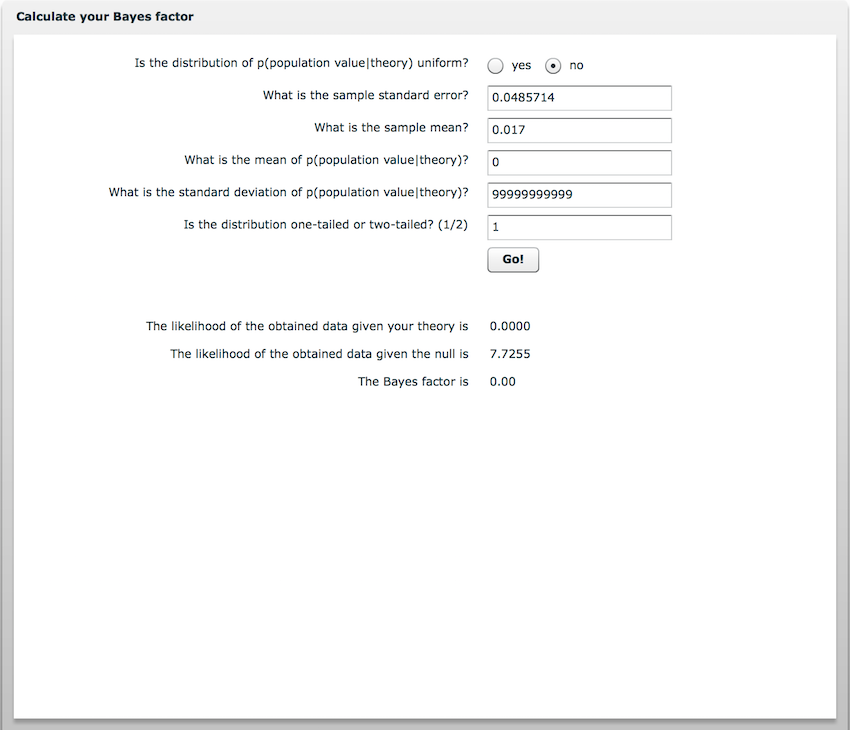


Figure 3. Martins et al, Maximum Robustness Region

This can be reported as ∞

The infinity symbol indicates that as the predicted effect size increased, the Bayes factor moved towards zero (in truth, the Bayes factor will never be exactly zero, but the calculator rounds down).

Lakens et al (2018) also tested a more informed model of H1, based on the results of a previous study conducted by Scheibe, Sheppes and Staudinger (2015).

Calculate the result that Scheibe et al obtained:

0.405 - 0.485 # young - old

## [1] -0.08

It is important that this is calculated the same way as we calculated the result of Martins et al (i.e., by subtracting old from young).

Scheibe et al (2015) obtain a mean difference in the opposite direction; they obtain a negative value, whereas Martins et al. obtained a positive value. You must therefore enter the obtained mean difference as a negative value when entering it in to the calculator. It is always the obtained mean difference that is entered as a negative value, because standard deviations cannot be specified with a negative value. Figures 4-6 demonstrate how to calculate the Bayes factors and robustness regions for this more specific model of H1:

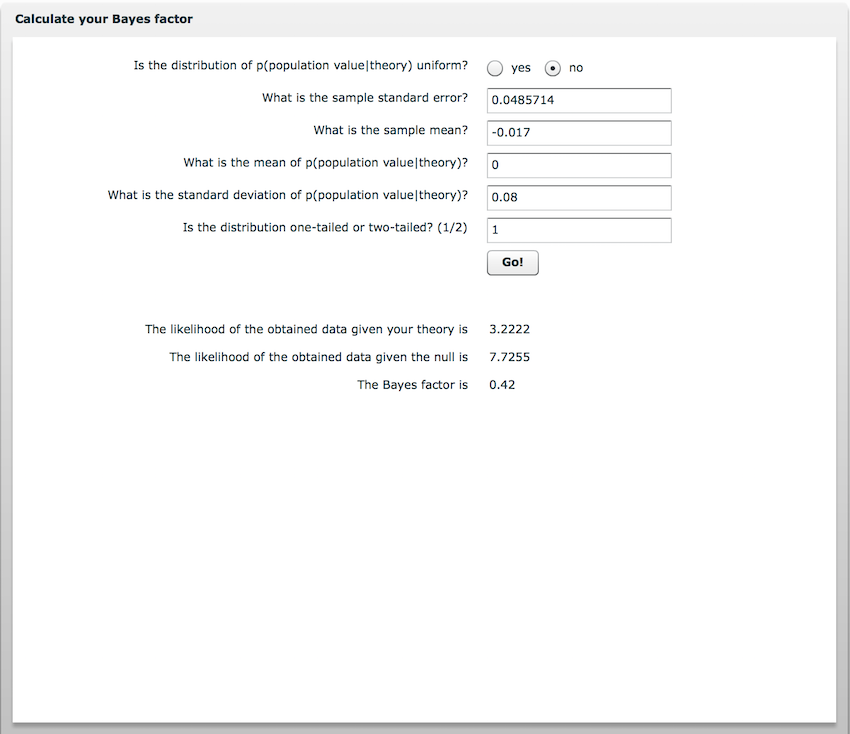


Figure 4. Martins et al, Bayes Factor (Informed)

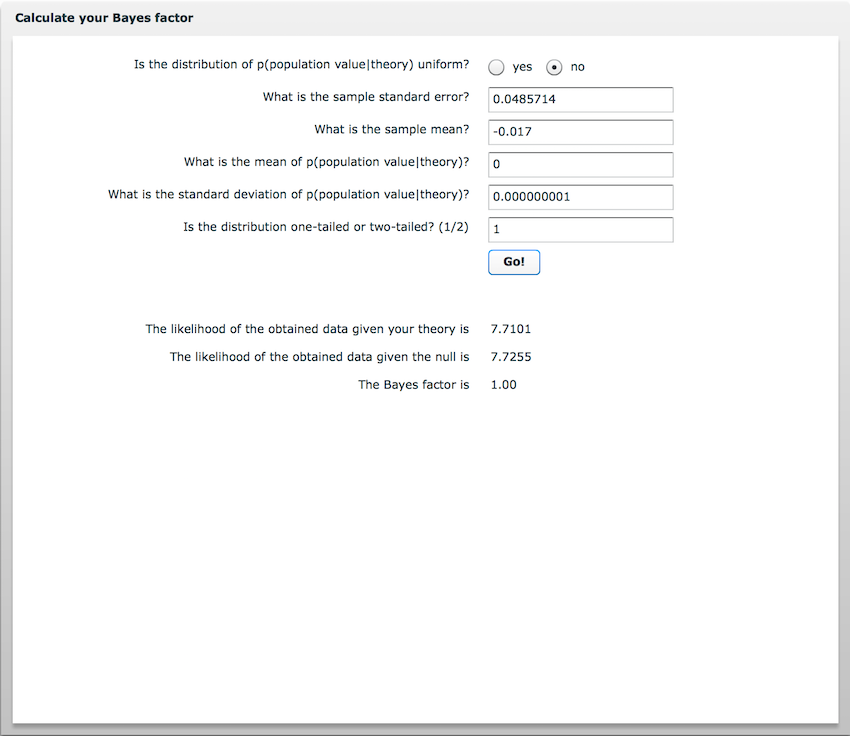


Figure 5. Martins et al, Minimum Robustness Region

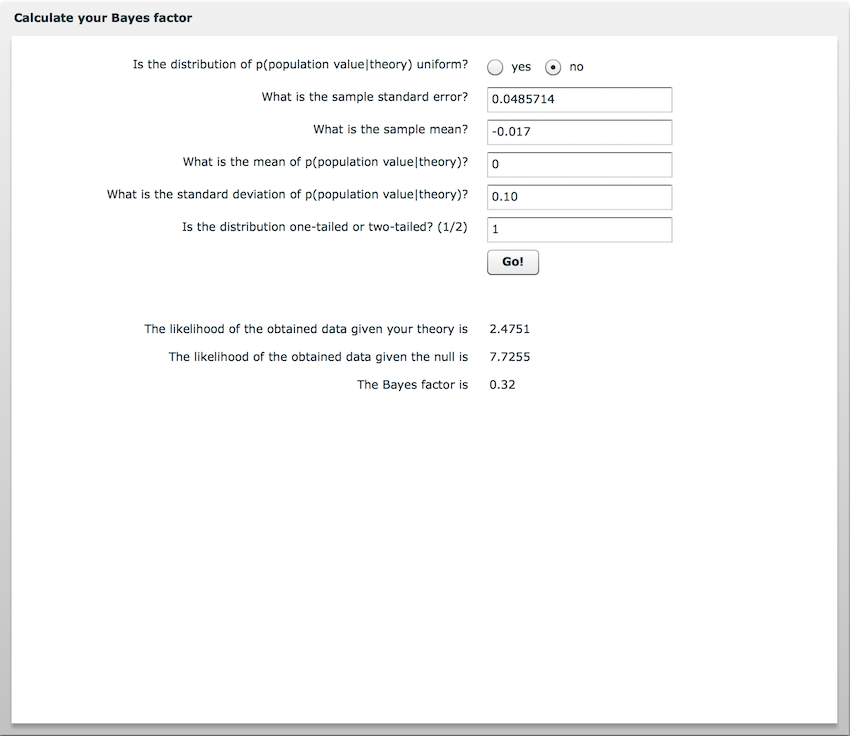


Figure 6. Martins et al, Maximum Robustness Region

To be reported as: .

## EXAMPLE 2: Shega, Tiedt, Grant and Dale (2014)

Shega et al. (2014) examined the relationship between pain and age. They compared three groups. Here we calculate a Bayes factor for each of the three comparisons.  
The process is identical to the previous example: calculate the mean difference and then divide it by t in order to derive the standard error.

In this example, though, the authors did not report t-tests, so the t-tests have been calculated using the TOSTER R package which calculates Welch’s t-test using the Ms, SDs, and Ns for each condition. Alternatively, you could use an online calculator to calculate the t-test (e.g., <https://www.graphpad.com/quickcalcs/ttest1/?Format=SD> - be sure to select ‘Welch’s Unpaired t test’, which is appropriate in this example because the number of participants differ across conditions).

The authors did not have a directional hypothesis. That is, they didn’t have a prediction about whether adults aged 62-69 would be in more or less pain than adults aged 70-79. This nondirectional hypothesis can modelled using a two-tailed normal distribution (see Figures 7 - 15). By specifying that the mean of our prediction (third box down) is zero, we center the model on zero, such that small effects closre to zero are considered more plausible than larger effects that are further away from zero.

The approximate range of effect (specified in the second box from the bottom of Figures 7 - 15) is based on the finding that 10-20% reductions on pain have been shown to be “noticeable”, and reductions up to 40% shown to be “meaningful” (Dworkin et al., 2008). Twenty percent on a 7-point Likert scale corresponds to a change of 1.21. Using this as an approximate scale of effect thus predicts that raw effects up to 40% are plausible.

Figures 7-15 demonstrate how to calculate the Bayes factors and Robustness Regions for each of the three comparisons.

### Example 2a: Shega et al., 62-69 vs 70-79

2.03 - 1.98

## [1] 0.05

0.05 / 0.4925451 # (t-test calculated from TOSTER package)

## [1] 0.1015135

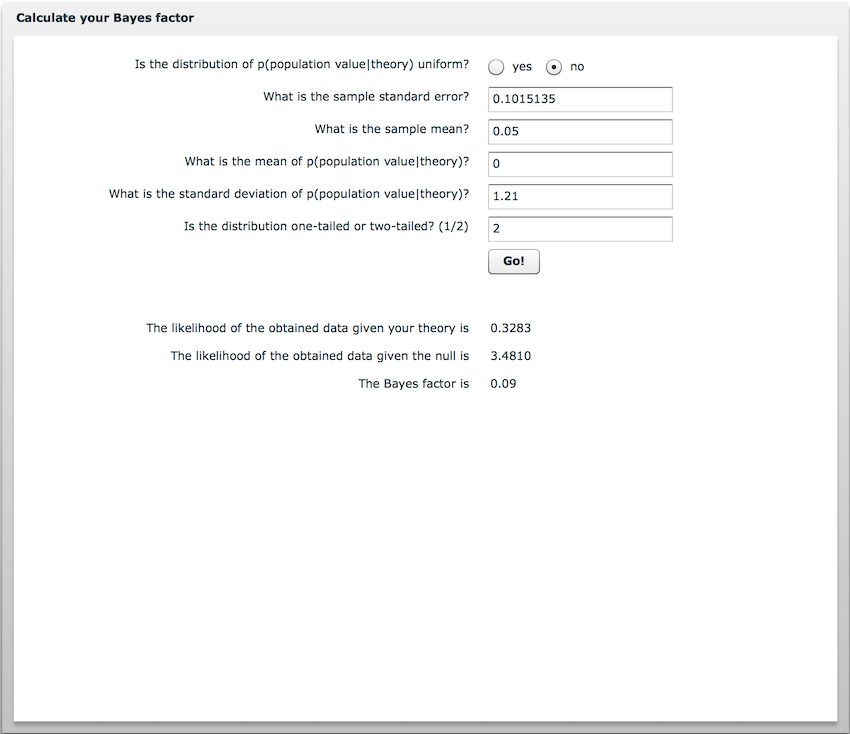


Figure 7. Example 2a: Shega et al., 62-69 vs 70-79 , Bayes Factor

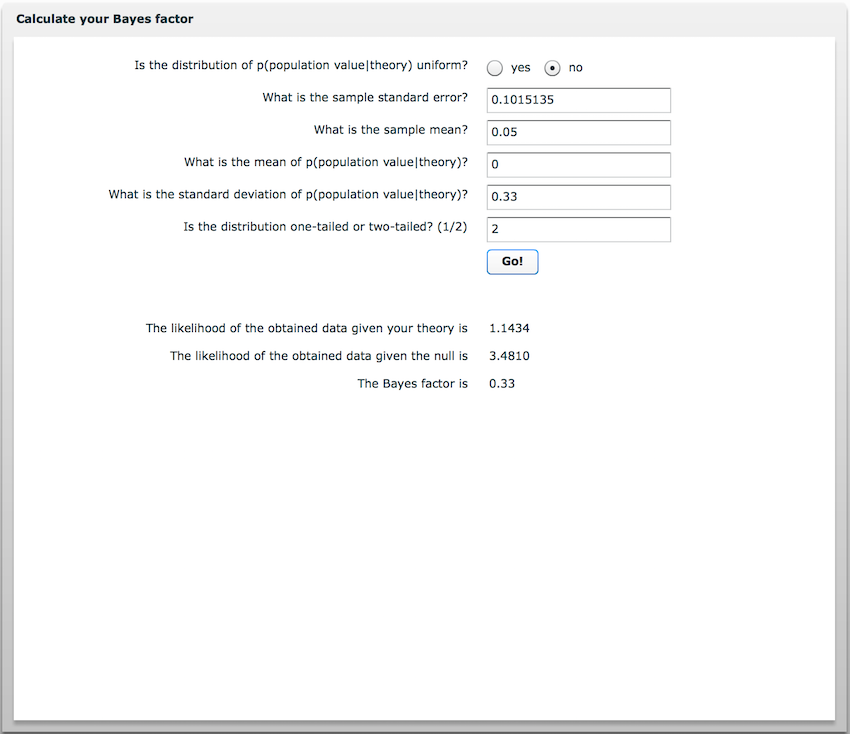


Figure 8. Example 2a: Shega et al., 62-69 vs 70-79, Minimum Robustness Region

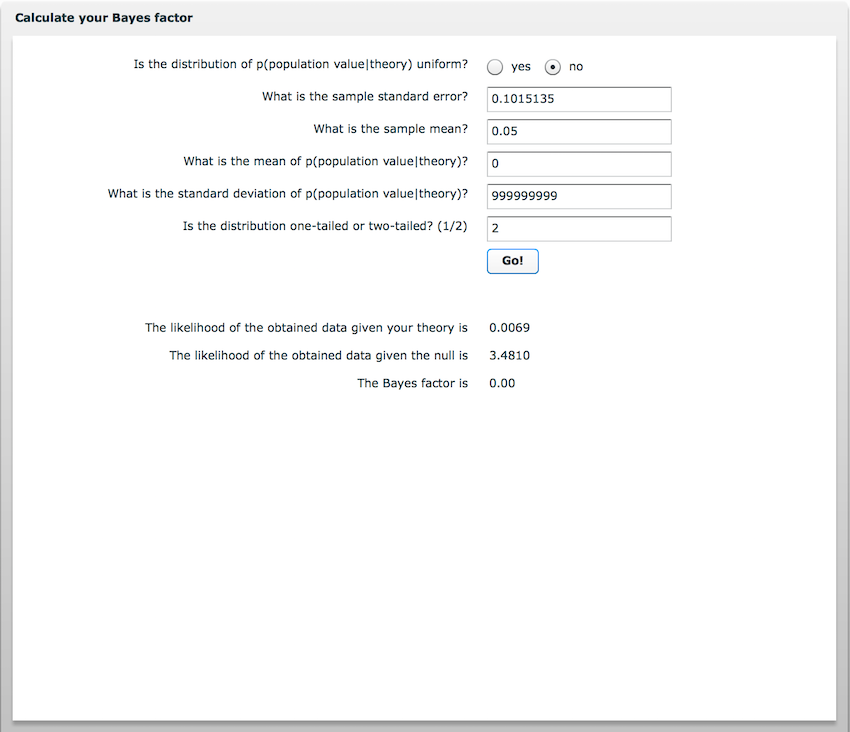


Figure 9. Example 2a: Shega et al., 62-69 vs 70-79, Maximum Robustness Region

Reported as:

### Example 2b: 62-69 vs 70-79

2.14 - 1.98 # obtained mean difference

## [1] 0.16

0.16 / 1.108473 # Standard Error (t-test calculated from TOSTER package)

## [1] 0.1443427

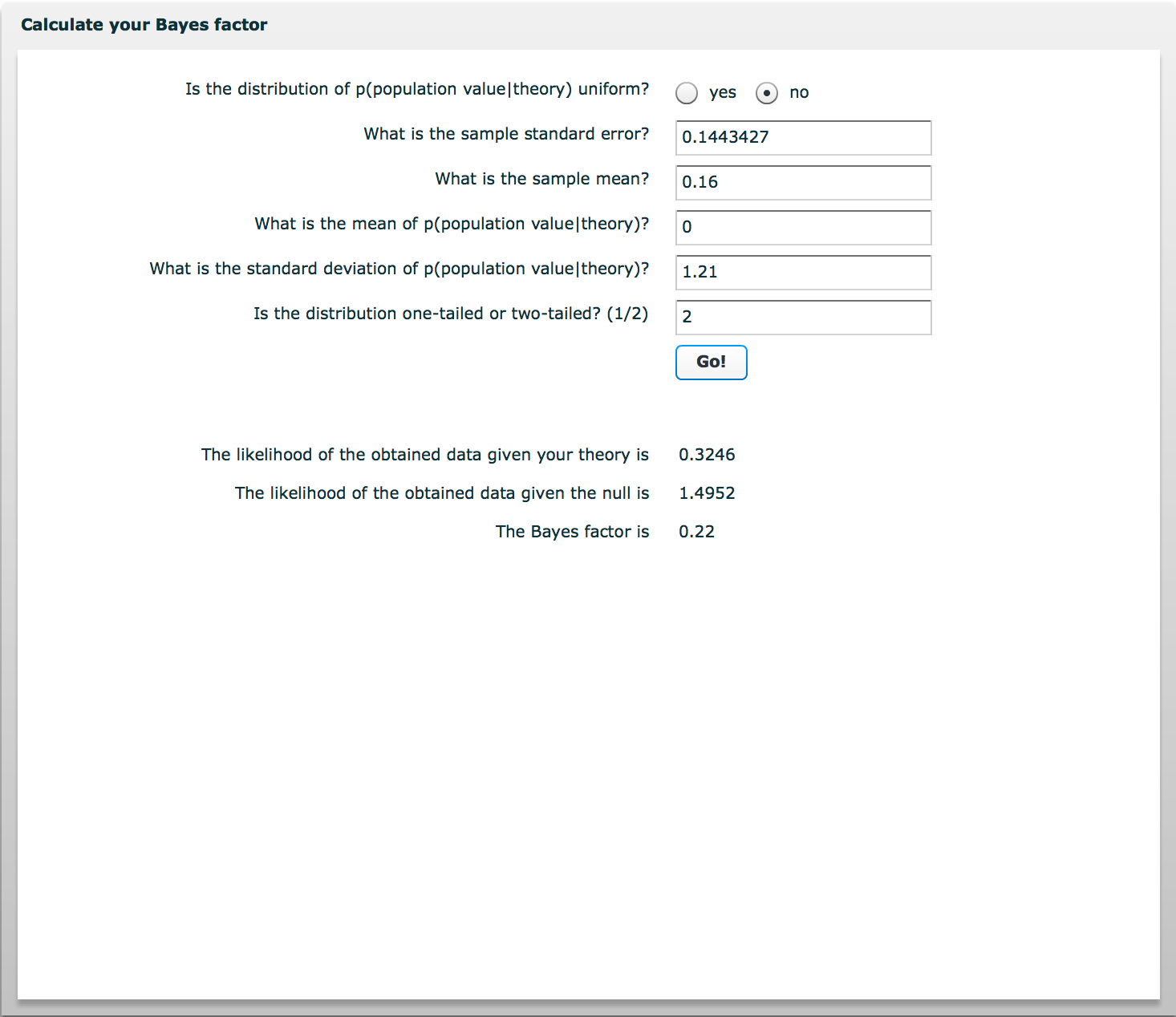


Figure 10. Example 2b: Shega et al., 70-79 vs 80+, Bayes Factor

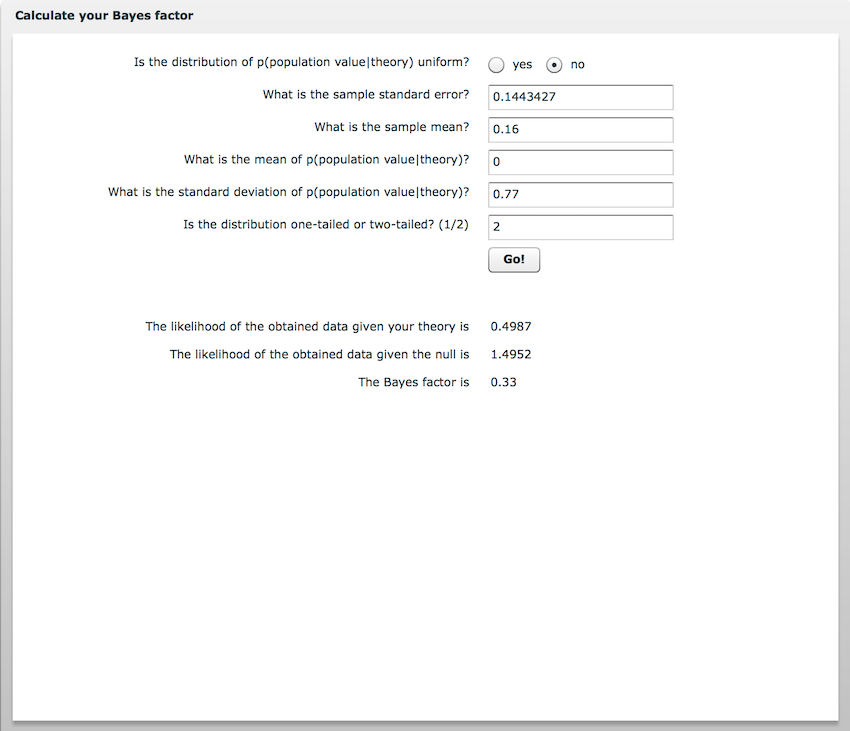


Figure 11. Example 2b: Shega et al., 70-79 vs 80+, Minimum Robustness Region

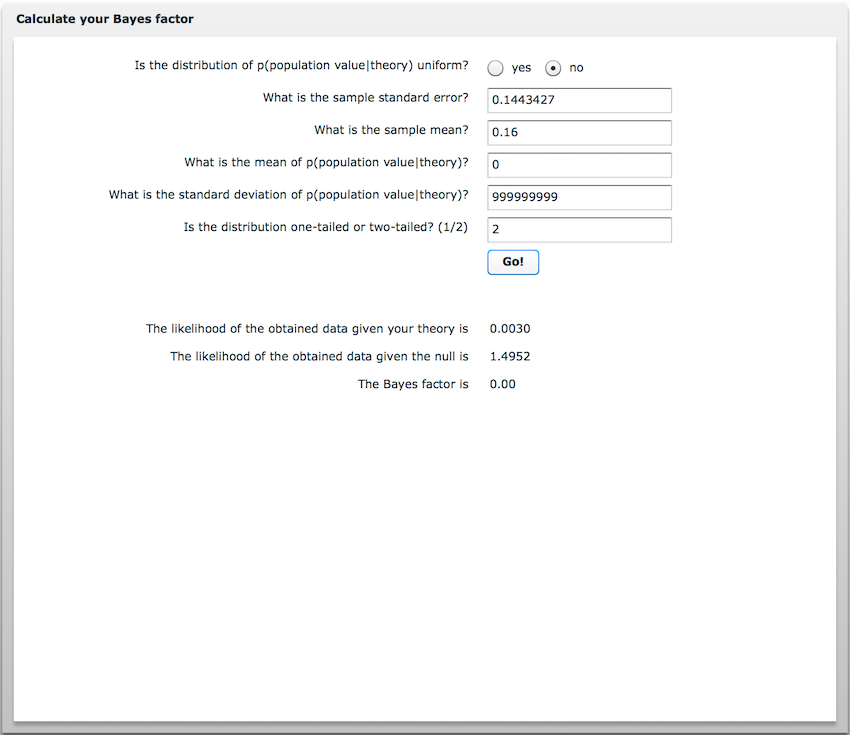


Figure 12. Example 2b: Shega et al., 70-79 vs 80+, Maximum Robustness Region

Reported as:

### Example 2c: 62-69 vs 70-79

2.14 - 2.03

## [1] 0.11

0.11 / 0.7255117 # (t-test calculated from TOSTER package)

## [1] 0.1516171

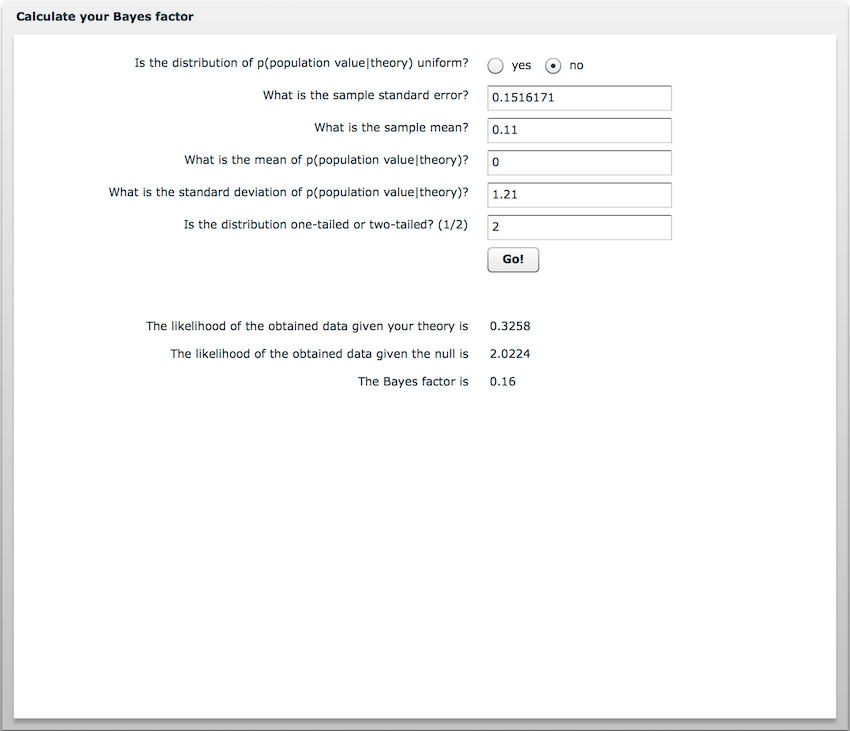


Figure 13. Example 2c: Shega et al., 62-69 vs 80+, Bayes Factor

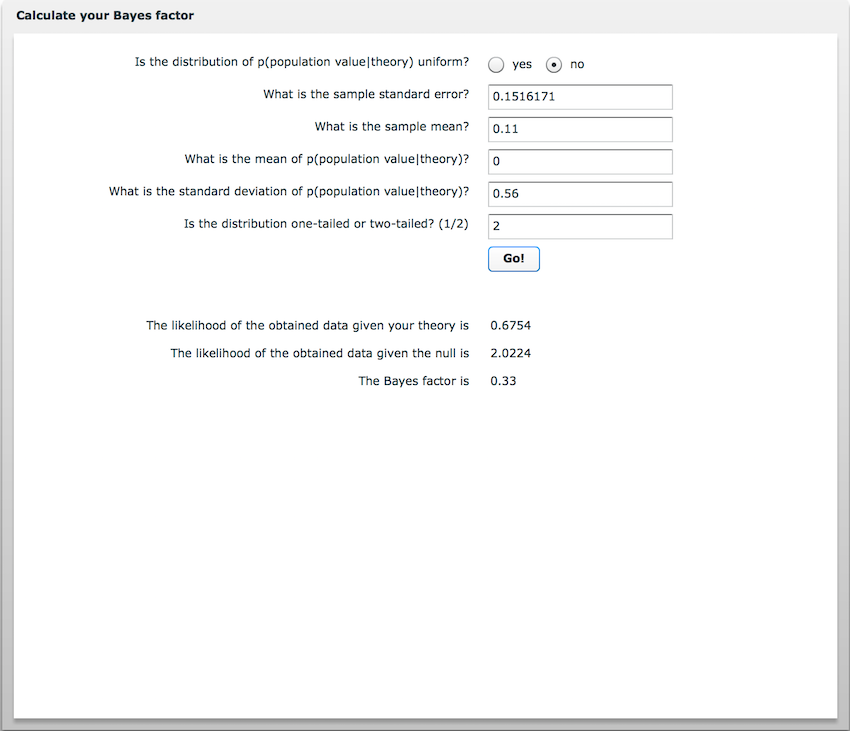


Figure 14. Example 2c: Shega et al., 62-69 vs 80+, Minimum Robustness Region

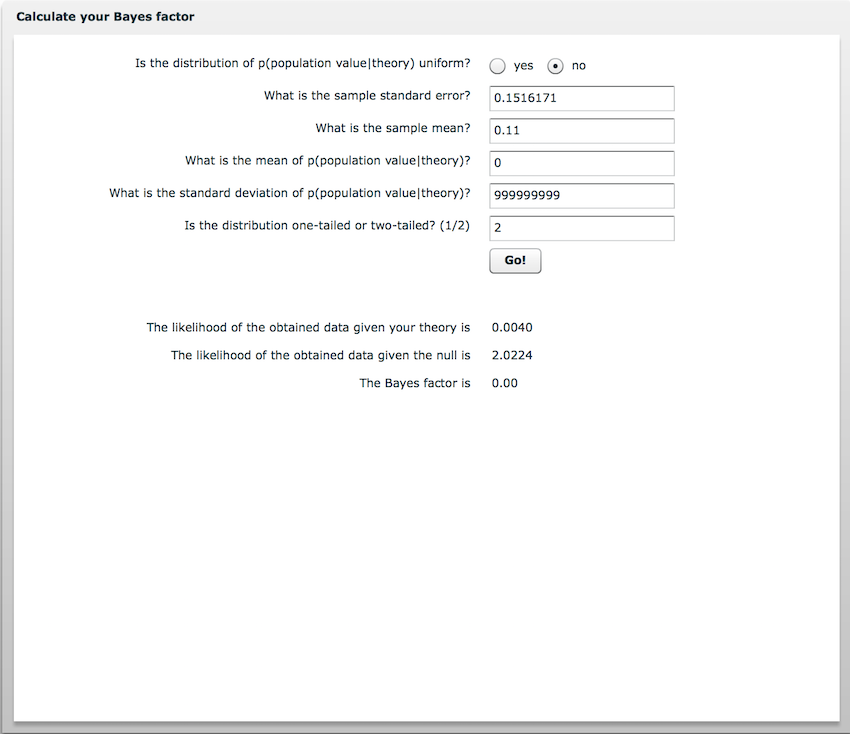


Figure 15. Example 2c: Shega et al., 62-69 vs 80+, Maximum Robustness Region

Reported as:

## EXAMPLE 3: Westerhof, Bohlmeijer and McAdams (2017)

Westerhof and colleagues obtained a non-significant correlation of r = .12 when comparing Openness with Despair. Does this constitute evidence for the null, evidence for H1, or is it inconclusive evidence?

Remember, Dienes’ (2008) calcualtor assumes normality, and r violates this assumption. Fortunately it is easy to calculate the raw beta coefficients if one knows the standard deviations of the two variables. Westerhof et al. helpfully provided these, and so we can calculate the raw beta coefficients as follows:

# Information re obtained data  
0.14 \* 0.8/0.6 # 0.19, prior beta

## [1] 0.1866667

0.12 \* 1.0/0.6 # 0.20, obtained beta

## [1] 0.2

To calculate the standard error of this raw beta coefficient, first normalise r using Fisher’s z transformation. When r <= .20, the transformation won’t change the result very much. Fisher’s z also has a known standard error: (1/ **√**(N-3)). Here I use the FishersZ() function available in the DescTools R package, but you could alternatively use an online calculator (e.g., <http://onlinestatbook.com/calculators/fisher_z.html>).

library(DescTools)

# Transform r to z  
FisherZ(.12)

## [1] 0.120581

# Calculate SE  
1/sqrt(218-3)

## [1] 0.06819943

Having calculated Fisher’s z and its SE, we can calculate a z-test:

0.12 / 0.068

## [1] 1.764706

This will be the same z-score for our raw beta coefficient: .20 / SE = 1.76. Thus the standard error for our raw beta coefficient is:

.20 / 1.76

## [1] 0.1136364

Hence, we have a raw beta coefficient = .20 which has a standard error of .114. As usual, these values are entered in to the first two boxes of the Dienes (2008) calculator (see Figure 16). To know whether this data provides evidence for the sort of effect size considered to be evidential elsewhere in the same paper (r = .14, corresponding to beta = .19), we can therefore model the alternative hypothesis as a half-normal scaled using a predicted effect of .19:

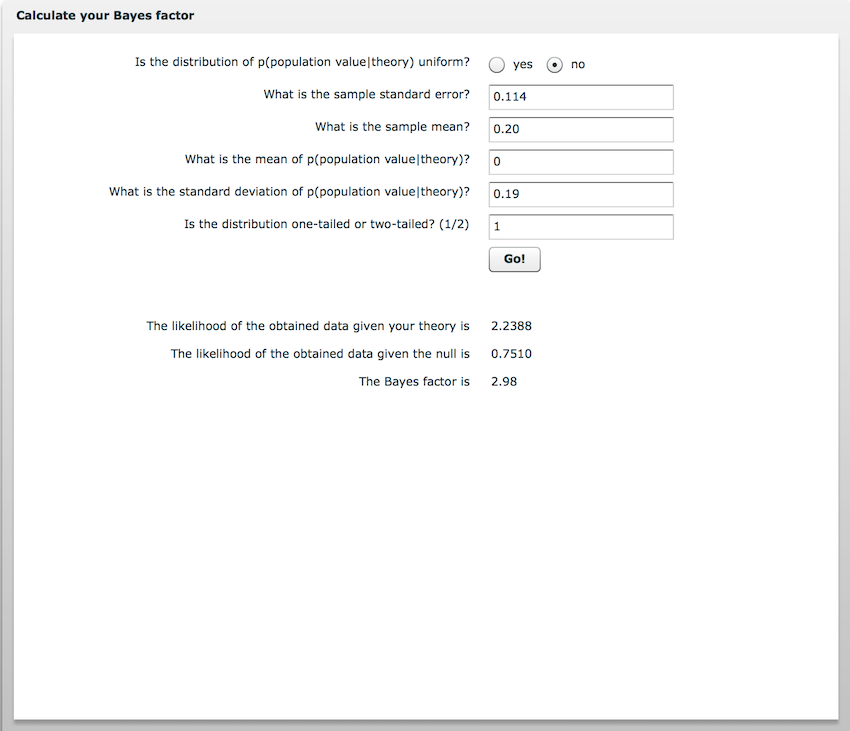


Figure 16. Example 3: Westerhof et al, Bayes factor

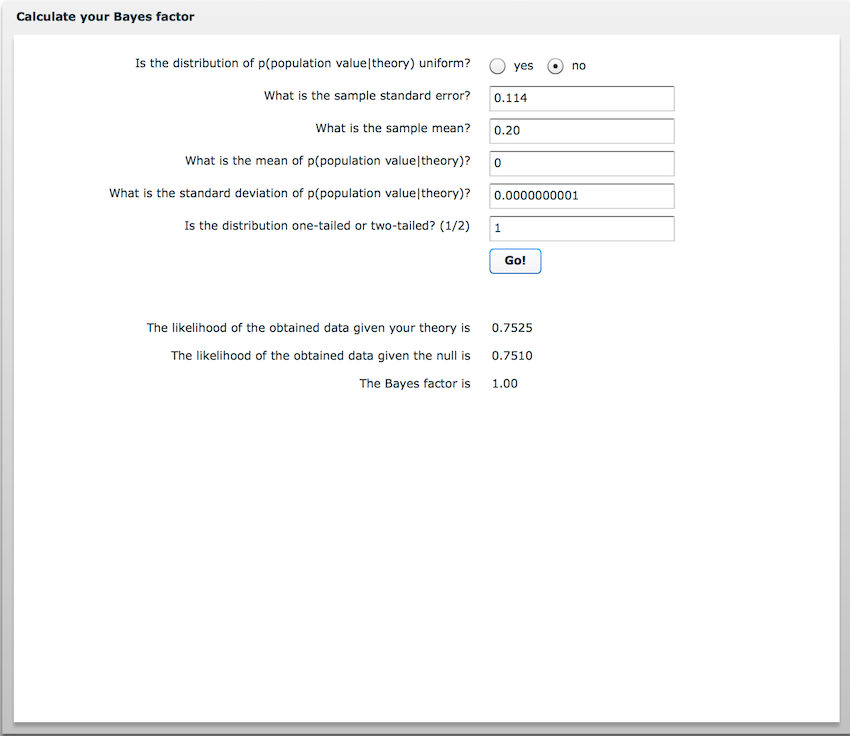


Figure 17. Example 3: Westerhof et al, Minimum Robustness Region

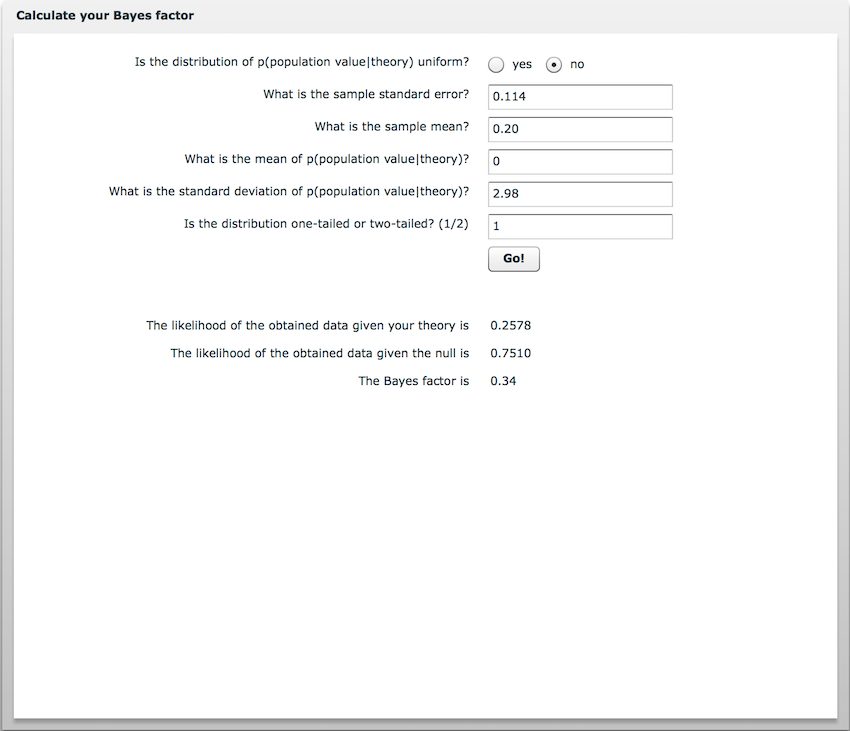


Figure 16. Example 3: Westerhof et al, Maximum Robustness Region

Reported as:

## 

## EXAMPLE 4: Spaniol, Schain and Bowen (2014)

Spaniol et al (2014) examined whether anticipating a reward would enhance long-term memory formation equally well in older and younger individuals. Here we calculate the interaction effect (the “difference of differences”) and it’s standard error:

(0.77-0.76) - (0.76-0.750001) # (YoungHigh-YoungLow) - (OldHigh-OldLow)

## [1] 1e-06

1e-06 / 5.088219e-05 (calculated from TOSTER function)

## [1] 0.01965324

We can specify the prior model using an effect size also reported by Spaniol et al. in their first experiment:

(0.61-0.54) - (0.64-0.61)

## [1] 0.04

We now have all the information needed to calculate a Bayes factor and its Robustness Regions (Figures 19-21).

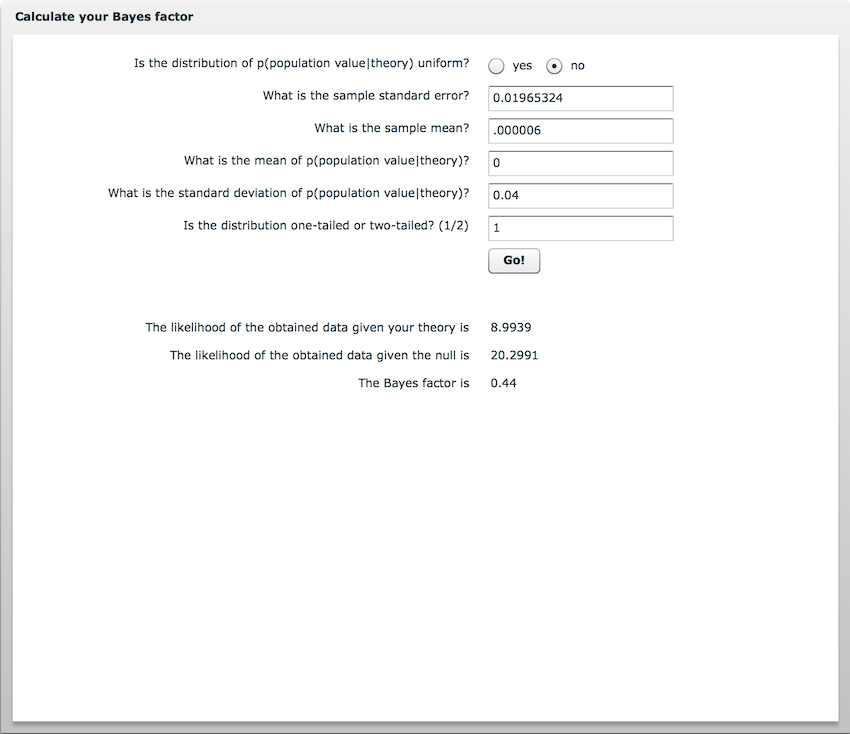


Figure 19. Example 4: Spaniol et al, Bayes Factor

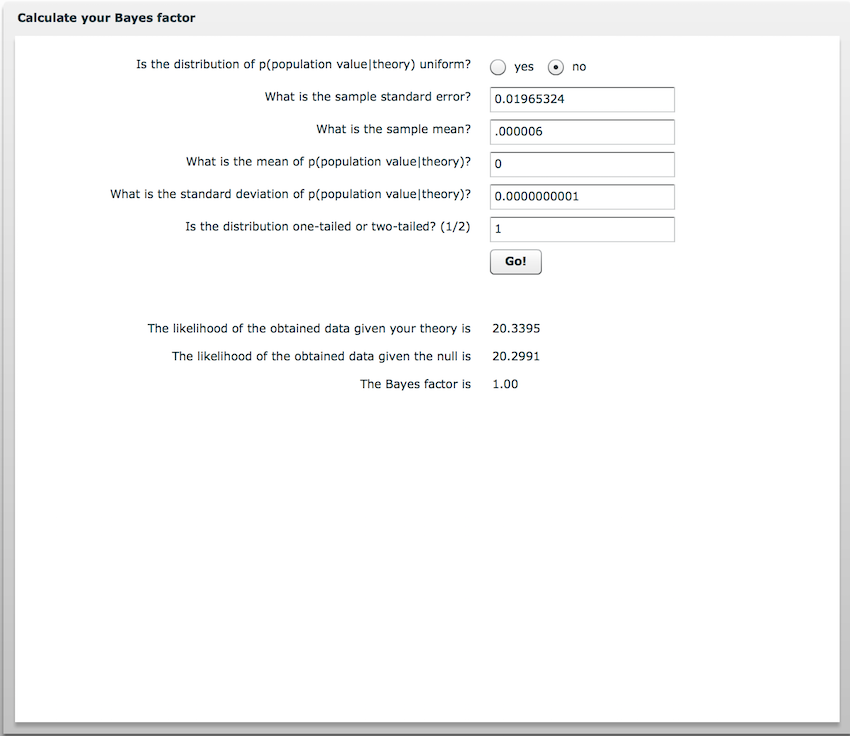


Figure 20. Example 4: Spaniol et al, Minimum Robustness Region

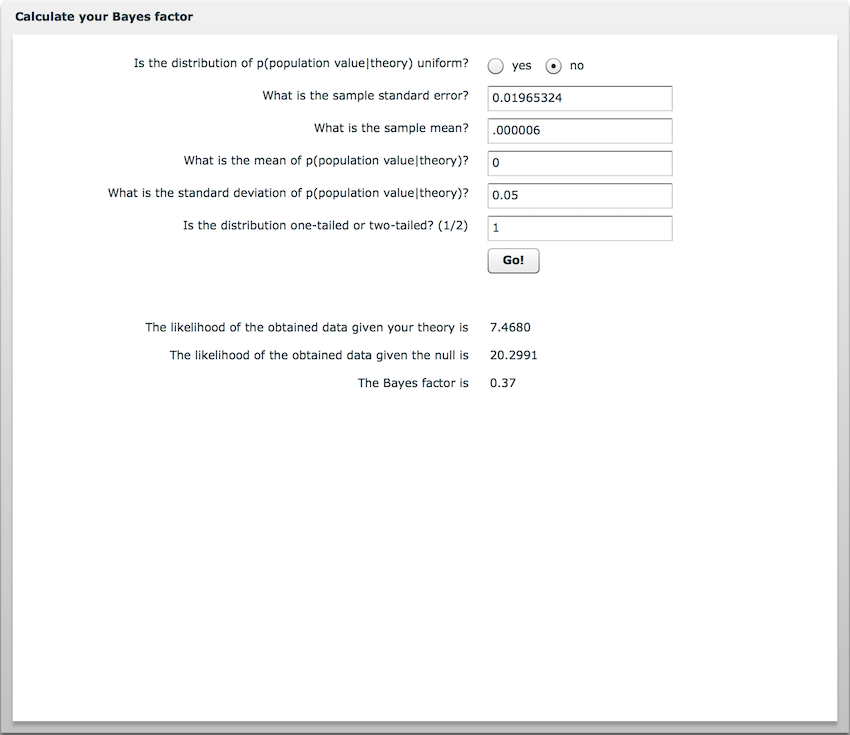


Figure 21. Example 4: Spaniol et al, Maximum Robustness Region

Reported as:

## 

## ESTIMATING SAMPLE SIZE

Instructional video available at: <https://www.youtube.com/watch?v=10Lsm_o_GRg>

Consider a researcher who aims to determine whether frail older adults demonstrate cognitive deficits relative to 11.5 nonfrail older adults. How many participants should the researcher expect to recruit in order to obtain conclusive evidence? Here, “conclusive evidence” is taken to mean B >= 3 or B <= 0.33.

The researcher could use the result of previous research, as well as its standard error, to determine exactly how many participants would need to be collected, first assuming H0 is true and then assuming H1 is true, before the evidence meets the threshold.

### Previous Research

Bunce, Batterham & Mackinnon (2018) reported that non-frail adults (M=11.53, SD=3.52, N=304) recalled significantly more names of animals than frail adults (M=10.11, SD=3.20, N=154), t(456) = 4.20, p < .001.

We can start by calculating the standard error implicitly reported by Bunce et al (2018):

bunce.se <- (11.53-10.11) / 4.20 # Calculated from Mdif / t

In order to calculate the standard error for different sample sizes, we will need to calculate the total number of participants recruited by Bunce et al (2018):

bunce.n <- 304+154 # 458

The predicted effect size for the replication is provided by the effect size reported by Bunce et al:

bunce.mdif <- 11.53-10.11 # 1.42  
obtain.h1 <- bunce.mdif

Notice we also create a variable called ‘obtain.h1’ which is identical to the effect size we are predicting. This is because we will shortly be calculating Bayes factors for varying levels of N, while assuming that our data obtained the exact predicted effec size. We will also be doing the same assuming for the null:

# Predicted effect size for H0 (here = 0)  
obtain.h0 <- 0

### Sample size estimation

The research needs to decide what the maximum number of participants they could recruit is. In this case, lets assume the researcher could recruit up to 250. We do this by creating a sequence of numbers ranging from 2-250 that increase by 1:

# Number of subjects we are willing to test  
n = seq(2, 250, 1)  
nmax=250-2 # 248

Next, we calculate the standard error for each of of these sample sizes. Notice that the standard error varies as a function of the square root of n, or in this case, the change in the magnitude of n:

se\_diff = (bunce.se) \* sqrt(bunce.n / n)

We can now calculate Bayes factors for each sample size ranging from 2 to 250 participants. Start by running the Bayes factor function from Dienes and McLatchie (2018):

## Run the Bayes factor t-distribution function from Dienes and McLatchie (2018)  
Bft<-function(sd, obtained, dfdata, meanoftheory, sdtheory, dftheory, tail = 2)  
{  
 area <- 0  
 normarea <- 0  
 theta <- meanoftheory - 10 \* sdtheory  
 incr <- sdtheory/200  
 for (A in -2000:2000){  
 theta <- theta + incr  
 dist\_theta <- dt((theta-meanoftheory)/sdtheory, df=dftheory)  
 if(identical(tail, 1)){  
 if (theta <= 0){  
 dist\_theta <- 0  
 } else {  
 dist\_theta <- dist\_theta \* 2  
 }  
 }  
 height <- dist\_theta \* dt((obtained-theta)/sd, df = dfdata)  
 area <- area + height \* incr  
 normarea <- normarea + dist\_theta\*incr  
 }  
 LikelihoodTheory <- area/normarea  
 Likelihoodnull <- dt(obtained/sd, df = dfdata)  
 BayesFactor <- LikelihoodTheory/Likelihoodnull  
 BayesFactor  
 }

First we calculate Bayes factors assuming that the mean difference obtained our our future study will be zero, thereby supporting the null hypothesis:

x <- 1/Bft(se\_diff , obtain.h0 , n-1, meanoftheory=0, sdtheory=bunce.mdif, dftheory=100000, tail=1)  
x

## [1] 1.068325 1.078144 1.092738 1.108393 1.124351 1.140358 1.156308  
## [8] 1.172153 1.187868 1.203438 1.218859 1.234127 1.249242 1.264206  
## [15] 1.279021 1.293688 1.308211 1.322593 1.336837 1.350945 1.364921  
## [22] 1.378769 1.392490 1.406088 1.419567 1.432927 1.446174 1.459308  
## [29] 1.472333 1.485251 1.498064 1.510776 1.523387 1.535901 1.548319  
## [36] 1.560644 1.572877 1.585020 1.597076 1.609045 1.620931 1.632733  
## [43] 1.644455 1.656098 1.667662 1.679151 1.690564 1.701905 1.713173  
## [50] 1.724370 1.735497 1.746557 1.757549 1.768476 1.779338 1.790136  
## [57] 1.800872 1.811546 1.822160 1.832714 1.843210 1.853649 1.864030  
## [64] 1.874356 1.884628 1.894845 1.905009 1.915121 1.925181 1.935191  
## [71] 1.945150 1.955060 1.964922 1.974735 1.984502 1.994222 2.003896  
## [78] 2.013525 2.023109 2.032650 2.042146 2.051601 2.061012 2.070383  
## [85] 2.079712 2.089000 2.098248 2.107457 2.116627 2.125758 2.134851  
## [92] 2.143906 2.152924 2.161906 2.170851 2.179760 2.188634 2.197473  
## [99] 2.206277 2.215048 2.223784 2.232487 2.241157 2.249794 2.258399  
## [106] 2.266972 2.275514 2.284024 2.292504 2.300952 2.309371 2.317760  
## [113] 2.326119 2.334449 2.342750 2.351022 2.359266 2.367482 2.375670  
## [120] 2.383830 2.391964 2.400070 2.408149 2.416203 2.424230 2.432231  
## [127] 2.440206 2.448156 2.456081 2.463980 2.471856 2.479706 2.487533  
## [134] 2.495335 2.503113 2.510868 2.518600 2.526308 2.533994 2.541656  
## [141] 2.549296 2.556914 2.564510 2.572083 2.579635 2.587165 2.594674  
## [148] 2.602162 2.609629 2.617074 2.624499 2.631904 2.639288 2.646652  
## [155] 2.653996 2.661320 2.668625 2.675910 2.683175 2.690422 2.697649  
## [162] 2.704857 2.712047 2.719218 2.726370 2.733505 2.740621 2.747719  
## [169] 2.754799 2.761861 2.768905 2.775933 2.782942 2.789935 2.796910  
## [176] 2.803868 2.810810 2.817735 2.824643 2.831534 2.838410 2.845268  
## [183] 2.852111 2.858938 2.865749 2.872544 2.879323 2.886087 2.892835  
## [190] 2.899568 2.906285 2.912988 2.919675 2.926347 2.933005 2.939647  
## [197] 2.946275 2.952889 2.959488 2.966073 2.972643 2.979199 2.985741  
## [204] 2.992269 2.998783 3.005284 3.011770 3.018243 3.024702 3.031148  
## [211] 3.037581 3.044000 3.050406 3.056799 3.063178 3.069545 3.075899  
## [218] 3.082240 3.088568 3.094884 3.101187 3.107477 3.113755 3.120021  
## [225] 3.126274 3.132516 3.138745 3.144962 3.151166 3.157359 3.163541  
## [232] 3.169710 3.175868 3.182014 3.188148 3.194271 3.200382 3.206482  
## [239] 3.212571 3.218648 3.224714 3.230769 3.236813 3.242846 3.248868  
## [246] 3.254879 3.260879 3.266868 3.272847

By counting the Bayes factors until we identify where B becomes greater than 3, we can identify the number of participants required. It is important to remember that first Bayes factor reported was calculated using 2 participants. Don’t forget to add an extra participant on after counting the Bayes factors.

In this case, the researcher would need 207 participants (highlighted in red) before they would obtain support for the null, assuming they find support for the null.

Now we can do the same for the alternative hypothesis:

y <- Bft(se\_diff , obtain.h1 , n-1, meanoftheory=0, sdtheory=bunce.mdif, dftheory=100000, tail=1)  
y

## [1] 1.045927 1.052427 1.062824 1.074370 1.086512 1.099066 1.111957  
## [8] 1.125144 1.138610 1.152343 1.166338 1.180592 1.195104 1.209874  
## [15] 1.224903 1.240194 1.255748 1.271568 1.287658 1.304019 1.320656  
## [22] 1.337571 1.354770 1.372255 1.390030 1.408101 1.426469 1.445142  
## [29] 1.464121 1.483413 1.503022 1.522953 1.543210 1.563799 1.584725  
## [36] 1.605993 1.627608 1.649576 1.671902 1.694593 1.717653 1.741089  
## [43] 1.764907 1.789113 1.813713 1.838714 1.864121 1.889942 1.916184  
## [50] 1.942853 1.969955 1.997500 2.025492 2.053941 2.082853 2.112235  
## [57] 2.142097 2.172445 2.203288 2.234634 2.266491 2.298868 2.331774  
## [64] 2.365216 2.399205 2.433748 2.468857 2.504539 2.540805 2.577664  
## [71] 2.615126 2.653202 2.691901 2.731235 2.771213 2.811847 2.853148  
## [78] 2.895126 2.937794 2.981163 3.025245 3.070051 3.115595 3.161889  
## [85] 3.208944 3.256775 3.305394 3.354815 3.405051 3.456117 3.508025  
## [92] 3.560791 3.614430 3.668955 3.724382 3.780727 3.838005 3.896232  
## [99] 3.955424 4.015598 4.076770 4.138957 4.202178 4.266448 4.331788  
##[106] 4.398214 4.465745 4.534400 4.604199 4.675162 4.747307 4.820655  
##[113] 4.895227 4.971045 5.048129 5.126500 5.206183 5.287197 5.369568  
##[120] 5.453318 5.538470 5.625049 5.713080 5.802587 5.893596 5.986133  
##[127] 6.080224 6.175896 6.273177 6.372093 6.472674 6.574948 6.678944  
##[134] 6.784692 6.892222 7.001566 7.112754 7.225819 7.340793 7.457709  
##[141] 7.576601 7.697502 7.820449 7.945475 8.072618 8.201914 8.333401  
##[148] 8.467116 8.603098 8.741387 8.882022 9.025045 9.170496 9.318419  
##[155] 9.468856 9.621851 9.777448 9.935694 10.096632 10.260312 10.426780  
##[162] 10.596086 10.768278 10.943407 11.121524 11.302682 11.486932 11.674331  
##[169] 11.864931 12.058790 12.255964 12.456511 12.660491 12.867962 13.078986  
##[176] 13.293626 13.511945 13.734006 13.959876 14.189621 14.423309 14.661009  
##[183] 14.902792 15.148728 15.398892 15.653356 15.912196 16.175490 16.443315  
##[190] 16.715750 16.992878 17.274779 17.561538 17.853241 18.149974 18.451825  
##[197] 18.758885 19.071244 19.388998 19.712239 20.041065 20.375574 20.715866  
##[204] 21.062042 21.414208 21.772467 22.136927 22.507698 22.884891 23.268619  
##[211] 23.658997 24.056142 24.460174 24.871215 25.289388 25.714818 26.147634  
##[218] 26.587967 27.035949 27.491715 27.955403 28.427153 28.907107 29.395410  
##[225] 29.892210 30.397657 30.911904 31.435107 31.967424 32.509017 33.060049  
##[232] 33.620688 34.191104 34.771471 35.361963 35.962761 36.574048 37.196009  
##[239] 37.828833 38.472714 39.127847 39.794431 40.472670 41.162771 41.864944  
##[246] 42.579404 43.306368 44.046059 44.798702

We see that we would require 82 participants (highlighted in red) before we could conclude there is sufficient evidence for the alternative hypothesis, assuming that we obtain results supporting the alternative hypothesis.

### Plotting the Sample Size Estimation

The following code will plot the estimation:

plot(x = x,  
 type = "l", # Specify that you want to plot a line graph  
 lwd = 2, # Thickness of line  
 lty = 2,  
 col = "black", # Colour of line  
 xlim=c(0,250), # Set limit of x-axis  
 ylim=c(0,10), # Set limit of y-axis  
 frame.plot=TRUE, # Do plot the frame of the graph  
 xlab="Sample Size", # Title for x-axis  
 ylab= "Bayes Factor", # Title for y-axis  
 axes = FALSE, # Don't plot the scales by default  
 main=paste("Estimating Sample Size for Bayes Factors")) # Main title  
segments(0, 3, 250, 3, col= 'black', lwd=2, lty=1) # Add a black horizontal line  
lines(y[y<10 ], lwd=2, lty=3)  
Axis(side=1,at=seq(0, 250, by = 10))  
Axis(side=2,at=seq(0,10, by=1))  
legend(0,10,legend=c("B for H0 (Mdiff = 0)","B for H1 (Mdiff = 1.42)", "Threshold for sensitive test (B = 3)"),  
 text.col="black", lty = c(2,3,1))

