

Bayes factor worksheet

Don't worry if you can't complete everything. Some of the material here is adapted from Chapter 11 of my book:

Baguley, T. (2012). *Serious stats: A guide to advanced statistics for the behavioral sciences*. Palgrave Macmillan.

Alternatively there is similar coverage in:

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan.

However, much of the material is inspired by work by Rouder, Morey and colleagues – notably the BayesFactor package and support material for it:

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes factors for common designs*. R package version 0.9.11. <http://CRAN.R-project.org/package=BayesFactor>

Installing the BayesFactor package

Installing a new package in R Studio is fairly easy. First locate the “Packages” tab (usually in the bottom right window where the “Plots” tab is). Then select “Install” and start typing the name of the package you want). Select this package from the list and click the “Install” button. You may be prompted to restart R as part of the process (so save any work in progress first). Once installed, you can load the package in the usual ways (e.g., from its name in the “Packages” window of R Studio).

For support with BayesFactor take a look at the online manual and blog:

<http://bayesfactorpcl.r-forge.r-project.org>

<http://bayesfactor.blogspot.co.uk>

Reprise: Effect of a single exposure to a US flag on voting intention

Q1

a) Use the `unit.prior.Bf.2s()` function to reproduce the “*Single exposure to a US flag example*” Bayes factor. First download the functions from

<http://www2.ntu psychology.net/seriousstats/SeriousStatsAllfunctions.txt>¹

The summary statistics for the effect of the flag on voting intention are: $M_1 - M_2 = 0.142$,

¹ You can also use load the functions directly into R using the `source()` function:
`source('http://www2.ntu psychology.net/seriousstats/SeriousStatsAllfunctions.txt')`

$t(181) = 2.02$, $SE = 0.0703$, $p < .04$. The call to the function takes the form:

```
unit.prior.Bf.2s(t, n1, n2, scale.factor=1)
```

Write down the call (i.e., including all the numbers) here:

b) The unit-information prior requires the scale factor is set to $r = 1$. A more general scaled-information prior can be obtained by increasing or decreasing r . *What do you think will happen to the Bayes factor if you increase r ?*

c) *What do you think will happen to the Bayes factor if you decrease r ?*

d) Re-run the call with $r = 0.5$ and $r = 2$. *Did the Bayes factor change in the way you expected?*

e) Now run it again with $r = 0.05$ or 0.01 (or both). *Did the results surprise you in any way? Why do think the Bayes factor behaved in this way as you went from $r = 2$ down to $r = 1$, $r = 0.5$ to $r = 0.05$ or $r = 0.01$?*

Q2

a) Use the `ttest.tstat()` function to get a JZS Bayes factor from the `BayesFactor` package. (The package is designed to work from raw data so this bit is not very elegant). The call to the function takes the form:

```
log.bf <- ttest.tstat(t, n1, n2, rscale='medium')[['bf']]
exp(log.bf)
```

... or (if you prefer) it in one step:

```
exp(ttest.tstat(t, n1, n2, rscale='medium')[['bf']])
```

Write down the call (i.e., including all the numbers) here:

b) The `ttest.tstat()` is an internal function not normally called directly by users of the package. *Why do you think it returns the (natural) logarithm of the Bayes factor rather than the Bayes factor itself?*

c) Now try and reproduce the test of the Bayes factor test of the directional hypothesis. The call to the function is identical but requires an extra argument:

```
nullInterval=c(0,Inf))
```

d) *What do you think will happen if the argument were `nullInterval=c(-Inf,0)` ? Try it. Does the output match what you expected?*

e) What do you think will happen if the argument were `nullInterval=c(-Inf, Inf)` ? Try it. Does the output match what you expected?

f) The `-Inf` or `Inf` (the R primitive for infinity) term in the `nullInterval` argument leaves that half of the prior density unconstrained, while 0 places no weight on that half of the prior density. This is perhaps an implausible because it is comparing a one-sided hypothesis (that δ is greater than/less than zero to a point null hypothesis – that $\delta = 0$ exactly).

An alternative approach (if the point null is implausible) is to compare the hypothesis $\delta > 0$ with the hypothesis $\delta < 0$. This is relatively simple because odds ratios like Bayes factors are multiplicative. Thus the ratio of the Bayes factor for $\delta > 0$ to the ratio of the Bayes factor $\delta < 0$ is the Bayes factor testing these competing hypotheses.

What is this Bayes factor?

Extended example using raw data

Wenzel (2004) carried out a small experiment investigating the effect of different types of diagrams on learning from text. Subsequent examples will refer to this data set as the diagram data. She gave 40 participants a short, four-paragraph text description of how SCUBA diving equipment works. Participants were assigned at random to one of four conditions:

text (presenting the four paragraphs on separate pages),

picture (presenting the text alongside a picture of a SCUBA diver),

full diagram (presenting the text alongside a labelled diagram depicting SCUBA equipment)

segmented diagram (presenting a relevant part of the full diagram on each of the four pages).

The principal hypotheses were:

1) that diagrams (but not pictures) would facilitate learning, and

2) that learning would be better for segmented diagrams than for the full diagram.

After a short delay, learning was measured by asking participants to write a brief explanation of how the SCUBA equipment worked. This was scored for description quality (using a method called propositional analysis). This produced scores ranging from 7 to 38 in the sample.

Q3 JZS *t* tests from data

a) You may already have the diagram data set in your R workspace. If not you can access it as follows (if you have a live internet connection):

```
diag.data <- read.csv("http://www2.ntupsychology.net/seriousstats/diagram.csv")
```

This data file contains five variables: *group* (a categorical factor), *descript* (the description quality DV), *time* (the time taken to read the text) and *condition* (a non-categorical numeric coding of the *group* variable).

To illustrate a two sample *t* test we want just two groups (control and segmented picture).

For convenience we'll do this by creating a vector of observations for each group:

```
cont <- diag.data$descript[1:10]
```

```
seg <- diag.data$descript[31:40]
```

A frequentist *t* test could be run as:

```
t.test(seg, cont)
```

The JZS BayesFactor *t* test takes the form:

```
ttestBF(seg, cont)
```

N.B. In both tests it is often convenient to put the group with the larger mean first so that the effect is positive (as this makes directional tests easier to interpret).

What is the JZS Bayes factor comparing the two conditions?

b) In this case a directional test is more plausible than no effect. Run a directional version of the analysis whether the segmented diagram aids learning from text. *What is the Bayes factor?*

c) *Did you use a point null hypothesis as the null hypothesis or did you compare the hypothesis that the segmented diagram improves learning vs. impedes learning? Does it matter?*

d) The approach can be extended to more complex designs. Use the function `anovaBF()` to run a one-way ANOVA on the `diag.data`. For a one-way design it has the format:

```
anovaBF(DV ~ factor, data)
```

Write down the full call to R and the Bayes factor it returns.

Note. It is often useful to assign the Bayes factor output to a new object (e.g., `bf.anov1`).

You can compare two models (under certain reasonable conditions such as one being nested in the other) by entering:

```
bf.mod1 / bf.mod2
```

This works because `/` is defined as a function to work on Bayes factor objects (in this package at least).

Q4 Correlation, regression and ANCOVA example

The JZS Bayes factor approach has been extended to a wide range of models including correlation, ANOVA, ANCOVA, multiple linear regression, and multilevel (mixed effect) models. (The latter includes repeated measures ANOVA as a special case.)

This example will focus on correlation, regression and ANCOVA.

a) Use the function `regressionBF()` to regress *descript* on *time* to see if the description quality is related to the time taken to complete a description.

```
regressionBF(descript ~ time, data = diag.data, rscaleCont = 'medium')
```


Write down Bayes factor it returns.

b) There isn't a separate correlation function in BayesFactor. (Hint you don't need to use R for this). *Write down the equivalent Bayes factor for the correlation.*

c) To combine categorical and continuous predictors (as in ANCOVA) we need to use the `lmBF()` function. Use the new data frame with *group* and *logtime* as predictors. The format for a simple ANCOVA would be:

```
lmBF(DV ~ covariate + factor, data)
```

Don't forget to keep the Bayes factor output by assigning to an object (e.g., `bf.ancov`)

Write down the full call to R and the Bayes factor it returns.

d) *How would you test the effect of group in the ANCOVA? What is the Bayes factor for this test?*

e) The omnibus *F* test in an ANCOVA or the equivalent 'omnibus' Bayes factor is not a particularly good test of the researcher's hypothesis. She actually predicted that the segmented diagram would be best, followed by the full diagram and that the picture and control conditions would be worst (with no preference in their order). The most appropriate test is therefore an *order-restricted* one. To test this hypothesis we need to impose both an *equality constraint* and an *order restriction*.

Start with the equality constraint. This requires you to re-run the ANCOVA model with *newgroup* rather than *group* as the factor. (Do this and keep this output in your workspace with the name `bf.ancov.eq`). Remember that creating *newgroup* just involved collapsing the two control groups (picture and text only) into a single control group.

What is the Bayes factor for the ANCOVA with the equality constraint?

f) Compare it to the ANCOVA model without equality constraint. *What is the Bayes factor?*

g) Now repeat the test of the differences in adjusted means in d), but using the equality constrained ANCOVA. *What is the Bayes factor?*

h) This final section uses order constraints with the ANCOVA. It may be helpful to read the BayesFactor package [blog entry on this](#) before going any further.

To check the order constraint in our equality restricted model we need to sample from the posterior distribution of the model with the equality constraint (the numerator of the Bayes factor). The vector `n.its` is used to set the number of samples taken.

```
n.its <- 50000
post.samples <- posterior(bf.ancov.eq, iterations = n.its, progress=TRUE)
```

We then need to find out how often the DV (*descript*) is in the predicted order: segmented > full > control.

```
consistent <- (post.samples[, "newgroup-segmented"] > post.samples[, "newgroup-
full"]) & (post.samples[, "newgroup-full"] > post.samples[, "newgroup-control"])
N_consistent <- sum(consistent)
prob.consistent <- N_consistent / n.its
```

How much more likely is the order-restricted model than the model with only the equality constraint? We can think of this as an evidential boost from having a riskier (more complex) theory (or an evidential penalty if the data don't support the theory). As there are six possible ways the three groups in *newgroup* could be ordered², it is simply:

```
evidence.boost <- 6 * prob.consistent
evidence.boost
```

The final step then (because Bayes factors are multiplicative) is to multiply the Bayes factor for the equality restricted ANCOVA with a suitable comparator model (e.g., the full ANCOVA

² The relevant quantity here is k factorial ($k!$) where k is the number of groups.

model with or without equality constraint) by this evidence boost.

Choose a suitable comparator model. Calculate the relevant Bayes factor. *What hypothesis does this comparison test?*

Note. There is more than one relevant hypothesis here, but pick the one that you think matches the original predictions of the researcher.