

Intro to Bayes worksheet

Don't worry if you can't complete everything. Much 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.

The central example is also taken from:

Howard, G., Maxwell, S., & Fleming, K. (2000). The proof of the pudding: An illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. *Psychological Methods*, 5, 315-332.

Normal distribution with known variance

Using Bayesian data analysis to re-evaluate a standardized mean difference

Note. The workshop example was a slightly non-standard example because it looked at the reported 'effect size' for a published study. The rationale is that Cohen's d family statistics (standardized mean differences) have a known variance. However, it does add a distraction in: i) deciding which d family statistic is appropriate (here Hedges' g), ii) obtaining the correct standard error. To remove this distraction it makes sense to use the `smd.post()` function (which was written for this workshop). This version of the functions works only for a 2 independent group design and assumes the statistic is Hedges' g .

Q1

a) Use the `smd.post()` function to reproduce the “*Single exposure to a US flag example*”. First download the functions.¹ The summary statistics for the effect of the flag on voting intention are: $M_1 - M_2 = 0.142$, $t(181) = 2.02$, $SE = 0.0703$, $d = 0.298$. (In that example I chose a subjective prior of 0.03 for the standardized mean difference and 0.06 for its standard deviation). The call to the function takes the form:

```
smd.post(smd.obs, smd.prior, sigma.prior, n1, n2=n1, plot=FALSE)
```

¹ The functions are available at

<http://www2.ntupsychology.net/seriousstats/SeriousStatsAllfunctions.txt>

You can also use load the functions directly into R using the `source()` function:

```
> source('http://www2.ntupsychology.net/seriousstats/SeriousStatsAllfunctions.txt')
```

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

b) *How does the posterior change when you use a more diffuse prior (with a higher sigma)?*

c) *How does the posterior change when you increase n to 5000? What does this suggest about the posterior distribution when N is large?*

Another known variance example

Howard et al. (2000) describe a Bayesian analysis of the effects of a *Psychology of Healthy Lifestyles* course on student alcohol consumption. They compare the Bayesian analysis to meta-analysis and frequentist inference. The present example reproduces part of their analysis (and only for Study 1).

The outcome measure was the self-reported increase in alcoholic drinks per week over the duration of the course. The treatment group received 6 hours instruction on the potential adverse effects of alcohol consumption in young people. Data were also collected for a control group (enrolled on different courses). The treatment group ($M = 1.58$, $SD = 2.19$, $n = 13$) increased their alcohol consumption less than the control group ($M = 2.21$, $SD = 2.98$, $n = 36$). The difference in means was not statistically significant and reported as $t(47) = 0.70$, $SE = 0.90$, $p = .48$, 95% CI [-1.2, 2.5].

Howard et al. set up priors separately for each group and compared subjective priors for a 'confident optimist' and a 'confident pessimist' with $\sigma_{prior} = 0.5$. The confident optimist anticipates a mean increase of only 1.5 drinks in the treatment group compared to an increase of 5.5 in the control group. The confident pessimist anticipates both groups increase by 5.5 drinks per week. For a normally distributed prior this indicates that both the optimist and pessimist are around 95% certain that the increase will be within $\pm 2s$ (i.e., ± 1 drink per week) of the prior.

Q2

a) Calculate the posterior variance for the difference (increase) in drinks per week between the groups. (We suggest you use R as a calculator for this part, but you could also use a spreadsheet or a calculator. Equations are given in the Appendix). Note that this is the same for the optimistic and pessimistic prior.

Work out the down the posterior variance.

b) *What is the posterior mean for the optimistic prior?*

c) *What is the 95% posterior probability interval for the optimistic prior?*

d) *What is the posterior mean for the pessimistic prior?*

e) *What is the 95% posterior probability interval for the pessimistic prior?*

f) Check your above answers using the function `Bayes.norm.2s()` with a call of the following form (without small sample correction):

```
Bayes.norm.2s(M.diff, SD.1, SD.2, diff.prior, sigma.prior, n1, n2,
confidence = 0.95, ssc = FALSE, plot = FALSE)
```

Some more R functions

For unknown variance problems Berry (1995) suggests applying a small sample correction to $\hat{\sigma}$ when $n < 30$ before calculating $\hat{\sigma}_{\hat{\mu}}$. Although this is an approximate solution it is a pretty good one. The functions `Bayes.norm.1s()` and `Bayes.norm.2s()` from Baguley (2012) implement the standard known variance calculations and include the small-sample correction. (These functions should already have been downloaded earlier).

Appendix: Key equations for reference

Posterior variance:

$$\sigma_{post}^2 = \left(\frac{1}{\hat{\sigma}_{\hat{\mu}}^2} + \frac{1}{\sigma_{prior}^2} \right)^{-1}$$

Posterior mean:

$$\mu_{post} = \left(\frac{\sigma_{post}^2}{\hat{\sigma}_{\hat{\mu}}^2} \right) \hat{\mu} + \left(\frac{\sigma_{post}^2}{\sigma_{prior}^2} \right) \mu_{prior}$$

Posterior probability (credibility) interval:

$$\mu_{post} \pm 1.96 \times \sigma_{post}$$