Solution to Exercise 1

Department of Methodology and Statistics
Utrecht University

Bayesian Statistics, Week 7

A: Input your data.

The data can be specified or sourced in R :

```
# Specify the data
dat <- list("y.PE" = 58, "n.PE" = 141,
                        "y.PE" = 40, "n.PC" = 143)

# Or source the data
source("Exercise 1 - Data.txt")
```

B: Specify your model


```
# Bayesian Statistics - Exercise 1 Model file solution

# Cognitive behavioral therapy for PTSD: Is PE more effective than the baseline PC?

model{

# likelihood of the data
y.PE ~ dbin(theta.PE, n.PE)
y.PC ~ dbin(theta.PC, n.PC)

# prior distributions
theta.PE ~ dbeta(1, 1)
theta.PC ~ dbeta(1, 1)

# contrast
RR <- theta.PC / theta.PE


}
```

C: Generating initial values

This will happen automatically for this model.

D: Obtaining posterior samples.


```
# 1. load the rjags library
library(rjags)   # library used to access JAGS with R
# 2. model, data and chains specification :
model.def <- jags.model(
  file =  "Exercise 1 - Model solution C.txt",
    data = dat, n.chains = 2)
# 3. burn-in period :
update(object = model.def, n.iter = 1000)
# 4. obtain samples from the posterior distribution
#    of the parameters and monitor these :
parameters <- c("theta.PE", "theta.PC", "RR")
res <- coda.samples(
  model = model.def, variable.names = parameters,
  n.iter = 10000)
```

E: Inspecting convergence.

This is will be explained later in the course

F: Substantive interpretation.


```
Iterations = 2001:12000
Thinning interval = 1
Number of chains = 2
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

             Mean      SD  Naive SE Time-series SE
RR         0.6924 0.11640 0.0008231      0.0010841
theta.PC   0.2826 0.03720 0.0002630      0.0003425
theta.PE   0.4124 0.04129 0.0002920      0.0003693

2. Quantiles for each variable:

             2.5%    25%    50%    75%  97.5%
RR         0.4850 0.6112 0.6844 0.7648 0.9438
theta.PC   0.2124 0.2572 0.2818 0.3070 0.3583
theta.PE   0.3319 0.3843 0.4122 0.4402 0.4929
```

F: Substantive interpretation.

RR: post.mean $= 0.69$, 95% CCI $= [0.49, 0.94]$. The recovery chance is smaller for the control group (PC) (post.mean $= .28$) than for the cognitive behavioral therapy group (PE) (post.mean$=.41$).

Because the 95% posterior CCI of the RR lies below 1, we could say that we are 95% certain that PE therapy gives a higher chance of recovery than PC therapy.

G: Analytical results.

$$\hat{\theta}_{PE} = \frac{(1+58)}{(1+58)+(1+141-58)} = \frac{59}{143} = .41$$

$$\hat{\theta}_{PC} = \frac{(1+40)}{(1+40)+(1+143-40)} = \frac{41}{145} = .28$$

$$RR = \frac{.41}{.28} = 1.46$$

or

$$RR = \frac{.28}{.41} = .69$$

G: Analytical results.

Note that such analytical solutions are not possible for many other models! There may be small deviations (in the decimals) between the analytical results and the estimates from R. Why?

H: Evaluating historical data.

There is no wrong or right answer, but researchers have to give arguments for the relative importance they attach to different sources prior knowledge. E.g., non-comparable populations, flawed study design...

I: Specifying informative priors.

If we give Ronald's data 'full weight' and discard Jessica's data (see answer K if you chose Jessica's data here):

```
# Bayesian Statistics - Exercise 1 Model file solution

# Cognitive behavioral therapy for PTSD: Is PE more effective than the baseline PC?

model{

# likelihood of the data
y.PE ~ dbin(theta.PE, n.PE)
y.PC ~ dbin(theta.PC, n.PC)

# prior distributions
theta.PE ~ dbeta(41, 66)
theta.PC ~ dbeta(46, 86)

# contrast
RR <- theta.PC / theta.PE


}
```

J. Assessing prior influence.


```
Iterations = 2001:12000
Thinning interval = 1
Number of chains = 2
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

            Mean      SD  Naive SE Time-series SE
RR        0.7879 0.09340 0.0006604      0.0008185
theta.PC  0.3126 0.02783 0.0001968      0.0002429
theta.PE  0.3991 0.03098 0.0002190      0.0002818

2. Quantiles for each variable:

            2.5%    25%    50%    75%  97.5%
RR        0.6190 0.7228 0.7825 0.8469 0.9845
theta.PC  0.2595 0.2932 0.3122 0.3313 0.3681
theta.PE  0.3400 0.3780 0.3984 0.4197 0.4613
```

J. Assessing prior influence.

The 95% CCIs of the RR are smaller, and still do not include 1. Therefore, the RR is still in favor of PE, but less strongly than with uninformative priors.

K: Specifying informative priors.

If we give Jessica's data 'full weight' and discard Ronald's data (see answer J if you chose Ronald's data here):

```
# Bayesian Statistics - Exercise 1 Model file solution

# Cognitive behavioral therapy for PTSD: Is PE more effective than the baseline PC?

model{

# likelihood of the data
y.PE ~ dbin(theta.PE, n.PE)
y.PC ~ dbin(theta.PC, n.PC)

# prior distributions
theta.PE ~ dbeta(121, 126)
theta.PC ~ dbeta(81, 196)

# contrast
RR <- theta.PC / theta.PE


}
```

K. Assessing prior influence.


```
Iterations = 2001:12000
Thinning interval = 1
Number of chains = 2
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

              Mean      SD  Naive SE Time-series SE
RR        0.6252 0.05842 0.0004131      0.0005049
theta.PC 0.2877 0.02216 0.0001567      0.0001985
theta.PE 0.4615 0.02511 0.0001776      0.0002208

2. Quantiles for each variable:

           2.5%    25%    50%    75%  97.5%
RR        0.5167 0.5847 0.6229 0.6643 0.7439
theta.PC 0.2454 0.2724 0.2874 0.3027 0.3315
theta.PE 0.4134 0.4442 0.4612 0.4785 0.5113
```

K. Assessing prior influence.

The 95% CCIs of the RR are smaller, and still do not include 1. Therefore, the RR is still in favor of PE, stronger than with uninformative priors.