Exercises V

1. Failure times of two kinds of components were as follows ($n = 10$ observations for both):

```
list(x=c(0.6, 1.3, 1.0, 3.0, 1.0, 5.2, 4.6, 4.7, 2.8, 0.2),nx=10,
     y=c(2.9, 0.8, 0.4, 3.4, 0.3, 1.0, 2.3, 2.2, 0.5, 1.9),ny=10)
```

Assuming exponential models $x_i \sim \exp(\lambda_1)$, $y_i \sim \exp(\lambda_2)$, and Gamma(0.001,0.001)-prior for both λ_1, λ_2, write a BUGS model for computing the posterior distribution of λ-parameters, and compute posterior probability of the hypothesis that the expected failure time for x is longer than for y. Finally, compute posterior predictive distribution for the next failure times for x_{11} and y_{11} in BUGS, and compute the posterior probability that $x_{11} > y_{11}$. (You can either exploit sufficient statistics or individual data points in BUGS model code. It is better to set initial values manually, to avoid BUGS generating unrealistic values from the very wide prior).

2. Janne Ahonen and Jakub Janda shared the Four Hills Tournament (Vierschanzentournee, Keski-Euroopan mäkiviikot) championship in 2006. Both scored a total of 1081.5 points from four competitions. Before the tournament, both took part in four other competitions. Their scores from all eight competitions were

```
ahonen = c(299.7, 255.2, 281.7, 238.0,   270.9, 262.2, 255.4, 293.0)
janda  = c(238.7, 285.6, 287.1, 252.2,   262.6, 264.7, 263.2, 291.0)
```

Assuming a normal model $\mathrm{N}(\mu_i, \sigma_i^2)$ for the scores ($i = 1, 2$ to index Ahonen and Janda) and the uninformative prior $\pi(\mu_i, \tau_i) \propto 1/\tau_i$, a posterior density of model parameters can be obtained for both jumpers. Compute this with BUGS and provide 95% intervals for μ_i. Can we say which jumper was statistically better? To find out, formulate the hypothesis that Ahonen was better as $H_0 : \mu_1 > \mu_2$ and compute $P(H_0 \mid \text{data}) = P(\mu_1 > \mu_2 \mid \text{data})$ in BUGS. Note that you can write the prior approximately as $\pi(\mu_i) = \mathrm{U}(0, L)$ with some large L and $\pi(\tau_i) = \mathrm{Gamma}(0.001, 0.001)$ and it helps to give initial values to BUGS by yourself (otherwise BUGS will generate them from wide prior distributions which leads to unrealistic values. If initial values are too far from the range of posterior, BUGS might not start running). Optional extra problem: assuming all except the last scores are observed for both, what is the probability that after the last competition their total score difference in the Four Hills Tournament is less than 1 point? (In BUGS `abs()` calculates absolute value).

3. Run the BUGS code of the linear model for York Rainfall data, both with and without standardization of the variables x. Plot the scatter plot of the joint posterior distribution of β_1, β_2 in BUGS (`Inference -> Correlations -> Correlation Tool`-window, where you can type 'beta[1]' and 'beta[2]' in the dialog boxes after 'nodes'). Note the difference between the two approaches. Set hypothetical value `xnew=60` and predict `ynew` for this scenario. The predictive distribution for `ynew` should be same in both approaches. Finally, set `mu` in `Sample Monitor Tool` and run iterations after it is set, to record results for it. Then select from the menu `Inference -> Compare -> Comparison Tool`-window, and write `mu` after 'node', and `y` after 'other' and `x` after 'axis' and click `model fit` in the same window. You should get a graphical presentation of the estimated regression line and 95% confidence intervals for it, together with plotted data points. How many data points fall outside the 95% range? How many would you expect?

4. The following data represent 109 individuals with measured pulse (`pulse1`) and a repeated measurement (`pulse2`). Some of the individuals had been sitting and some running between the measurements (`status=1=run, status=2=sit`). Explain what model is coded in BUGS and what are the parameters there. Can we conclude that the runners have significantly higher 'pulse2' than no-runners? What can be said about the variability in 'pulse2' in both groups? (You can also plot pulse1 against pulse2 in R for both groups. Can you detect a possible fake-runner?).

```
model{
for(i in 1:2){
tau[i] ~ dgamma(0.01,0.01);
       var[i] <-1/tau[i]; sig[i]<-sqrt(var[i])
beta[i] ~ dnorm(0,0.001)
}
for(i in 1:N){
pulse2[i] ~ dnorm(mu[i],tau[status[i]])
mu[i]  <- beta[status[i]]*pulse1[i]
}
difference <- beta[1]-beta[2]
diff.is.positive <- step(difference)
}
# inits:
list(beta=c(1,1),tau=c(0.1,0.1))
# data:
list(N=109,
pulse1=c(
  86,  82,  96,  71,  90,  78,  68,  71,  68,  88,  76,  74,  70,  78,  69,  77,
  64,  80,  83,  78,  88,  70,  78,  80,  68,  70,  62,  81,  78,  86,  59,  68,
  75,  74,  60,  70,  80,  58,  84, 104,  66,  84,  65,  80,  66, 104,  76,  70,
  66,  92,  70,  63,  65,  76,  56,  64,  60,  68,  80,  65,  47,  50,  80,  76,
  70,  76,  72,  80,  76,  85,  49,  76, 145,  83,  72,  60,  80,  70,  68,  78,
  52,  74,  75,  72,  80,  84,  74,  90,  61,  85,  78,  76,  90,  64,  64,  88,
  64,  82,  88,  74,  88,  92,  76,  71, 119,  90,  86,  69,  75),
pulse2=c(
   88, 150, 176,  73,  88, 141,  72,  77,  68, 150,  88,  76,  71,  82,  67,  73,
   63, 146,  79,  79,  86,  98,  74,  76,  69,  96,  59,  79, 168, 150,  92, 125,
   130, 168, 104, 119, 140,  58,  84,  92,  68,  90,  67,  80,  60,  96,  76,  68,
   89,  84,  95,  65,  67,  74, 110, 126,  56,  84,  72,  82, 136,  90,  76,  72,
   74, 132, 115,  80, 150, 130,  83,  73, 155,  84, 136,  62,  82, 120, 136, 129,
   60,  72,  75,  68,  73, 140,  72, 160,  59, 131, 132,  80,  84,  68, 120, 144,
   64,  87, 120,  70, 136, 120, 168, 125, 120,  89,  84,  64, 68),
status=c(
  2, 1, 1, 2, 2, 1, 2, 2, 2, 1, 1, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 1, 2, 2, 2, 1, 2,
  2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 1, 2, 2, 2,
  1, 1, 2, 1, 2, 1, 1, 1, 2, 2, 2, 1, 1, 2, 1, 1, 1, 2, 1, 2, 1, 2, 2, 1, 1, 1, 2,
  2, 2, 2, 2, 1, 2, 1, 2, 1, 1, 2, 2, 2, 1, 1, 2, 2, 1, 2, 1, 1, 1, 1, 2, 2, 2, 2, 2)
  )
```