1. Use the stepwise DIC model selection procedure in BUGS for variable selection, starting with (1) the full model and (2) the null model. The response variable is oxygen uptake (O2UP) in milligrams of oxygen per minute. The set of covariates is biological oxygen demand (BOD), total Kjeldahl nitrogen (TKN), total solids (TS), total volatile solids (TVS), and chemical oxygen demand (COD). These can be collected into matrix X, where rows represent individuals, and columns represent covariates. The model is

$$\log(\text{O2UP}_i) \sim \text{N}(\mu_i, \sigma^2)$$
$$\mu_i = \beta_0 + \sum_{j=1}^{p} \gamma_j \beta_j X_{ij}$$

```
list(n=20, p=5, gamma=c(1,1,1,1,1),
BOD = c(1125, 920, 835, 1000, 1150, 990, 840, 650, 640, 583,
        570, 570, 510, 555, 460, 275, 510, 165, 244, 79),
TKN = c(232, 268, 271, 237, 192, 202, 184, 200, 180, 165,
       151, 171, 243, 147, 286, 198, 196, 210, 327, 334),
TS = c(7160, 8804, 8108, 6370, 6441, 5154, 5896, 5336, 5041, 5012,
      4825, 4391, 4320, 3709, 3969, 3558, 4361, 3301, 2964, 2777),
TVS = c(85.9, 86.5, 85.2, 83.8, 82.1, 79.2, 81.2, 80.6, 78.4, 79.3,
       78.7, 78, 72.3, 74.9, 74.4, 72.5, 57.7, 71.8, 72.5, 71.9),
COD = c(8905, 7388, 5348, 8056, 6960, 5690, 6932, 5400, 3177, 4461,
       3901, 5002, 4665, 4642, 4840, 4479, 4200, 3410, 3360, 2599),
O2UP = c(36, 7.9, 5.6, 5.2, 2, 2.3, 1.3, 1.3, 0.6, 0.7, 1, 1,
        0.8, 0.6, 0.4, 0.7, 0.6, 0.4, 0.3, 0.9))
```

Using the code below, start with some 'current model' (e.g. the full model, specified by gamma = c(1,1,1,1,1)). Compute DIC and check which candidate model is the best. Take the best model as the new 'current model' and write it as a new gamma=c(...). Compute again DIC and check which of the new candidate models is now best. Take the best as 'current' and keep repeating, until you find that the best model is the 'current'. Start the whole stepwise procedure from some other initial model and check if it gives the same final model choice in the end. (This is not guaranteed because the stepwise method is not globally searching among *all* models).

```
model{
  # ----------------------------------------
  # current model
  # ----------------------------------------
  # initial definition of variables for the current model
    for (i in 1:n){
          y[i] <- log(O2UP[i])
              x[i,1]<-BOD[i]
              x[i,2]<-TKN[i]
              x[i,3]<-TS[i]
              x[i,4]<-TVS[i]
              x[i,5]<-COD[i]
              }
  # model specificication of the current model
      for (i in 1:n){
```

```
            y[i] ~ dnorm( mu[i], tau )
            mu[i] <- beta0 + beta[1] * gamma[1]*x[i,1] +
                             beta[2] * gamma[2]*x[i,2] +
                             beta[3] * gamma[3]*x[i,3] +
                             beta[4] * gamma[4]*x[i,4] +
                             beta[5] * gamma[5]*x[i,5]
    }
# prior distributions
    beta0~dnorm(0, 0.001)
    for (j in 1:p){
            beta[j]~dnorm(0, 0.001)
    }
    tau~dgamma( 0.001, 0.001)

# ------------------------------------------------------------------------------
    # definition of gammas for candidate models
    for (k in 1:p){
            for (j in 1:p){
  gamma.can[j,k]<-gamma[j]*(1-equals(k,j))+(1-gamma[j])*equals(k,j)
                  }
    }
    # response data for candidate models
  for (i in 1:n){
            y1[i] <- y[i]
            y2[i] <- y[i]
            y3[i] <- y[i]
            y4[i] <- y[i]
            y5[i] <- y[i]
            y1[i] ~ dnorm( mu.can[i,1], tau.can[1] )
            y2[i] ~ dnorm( mu.can[i,2], tau.can[2] )
            y3[i] ~ dnorm( mu.can[i,3], tau.can[3] )
            y4[i] ~ dnorm( mu.can[i,4], tau.can[4] )
            y5[i] ~ dnorm( mu.can[i,5], tau.can[5] )
    }
#
# linear predictors for each model
    for (k in 1:p){
    for (i in 1:n){
            mu.can[i,k] <- beta0.can[k] +
            beta.can[1,k] * gamma.can[1,k]*x[i,1] +
            beta.can[2,k] * gamma.can[2,k]*x[i,2] +
            beta.can[3,k] * gamma.can[3,k]*x[i,3] +
            beta.can[4,k] * gamma.can[4,k]*x[i,4] +
            beta.can[5,k] * gamma.can[5,k]*x[i,5]
    }
    }
```

```
        #
        # prior distributions for parameters of candidate models
        for (k in 1:p){
            beta0.can[k]~dnorm(0, 0.001)
                for (j in 1:p){
                        beta.can[j,k]~dnorm(0, 0.001)
                }
                tau.can[k]~dgamma( 0.001, 0.001)
        }
 }
```

2. Use zeros trick to implement prior probability density $\pi(x) = -\log(x), x \in (0,1]$ and $\pi(y) =$ Cauchy$(0,1)$. Cauchy(μ,γ) density is $\gamma/[(\gamma^2 + (x-\mu)^2)\pi]$, where $\pi = 3.1415926\ldots$. For Cauchy you may monitor the median which is defined, unlike mean and variance which are not defined for Cauchy.

Below two ways of producing the distribution, first via zeros trick, then some other construction for comparison.

```
model{ # prior densities
# -log(x)-density:
zero <- 0
zero ~ dpois(L)
L <- -log(-log(x[1]))
x[1] ~ dunif(0,1)
x[2] ~ dunif(0,upper); upper ~ dunif(0,1)

# Cauchy density:
y0 <-0; g <- 1  # Cauchy parameters
pii <- 3.1415926; C<- 10000
# -log Cauchy density:
LL <-  -log(g)+log(pow((y[1]-y0),2)+pow(g,2)) +log(pii) + C
zeroo ~ dpois(LL)
zeroo <- 0
y[1] ~ dflat()
y[2] <- z1/z2; z1 ~ dnorm(0,1); z2 ~ dnorm(0,1)
# this will give Cauchy for y[2] but resulting from
# deterministic function of z1 and z2, instead of as a
# stochastic node for y[2].
}
list(y=c(1,NA),z1=1,z2=1)
```

3. Analyze the following data with Poisson log-linear model. Make box-plots of the age group effects and city effects. You may choose the first age group in the first city as a reference group. Obtain estimates of incidence per 100000 per year in each group of 'age'×'city'. Cases: The number of lung cancer cases 1968-1971 in Denmark. Pop: The population of each age group in each city. Age: The

age groups 1-6; of 40-54, 55-59, 60-64, 65-74 or >74. City: The city; one of Fredericia (1), Horsens (2), Kolding (3) or Vejle (4). Note that if Poisson intensity is scaled directly to 100000 per year as a parameter in the model, you may get numerical problems. Likewise, in similar problems with small incidence, the corresponding parameter is good to be rescaled.
http://www.sci.usq.edu.au/staff/dunn/Datasets/glms/poisson/danishlc.html

```
cases[] pop[]   age[]   city[]
11      3059    1       1
11      800     2       1
11      710     3       1
10      581     4       1
11      509     5       1
10      605     6       1
13      2879    1       2
6       1083    2       2
15      923     3       2
10      834     4       2
12      634     5       2
2       782     6       2
4       3142    1       3
8       1050    2       3
7       895     3       3
11      702     4       3
9       535     5       3
12      659     6       3
5       2520    1       4
7       878     2       4
10      839     3       4
14      631     4       4
8       539     5       4
7       619     6       4
END
```

```
model{
for(i in 1:24){
cases[i] ~ dpois(mu[i]); group[i] <- i
mu[i] <- pop[i]*4*lambda[i] # lambda = incidence per year
LA[i] <- lambda[i]/100000 # LA = incidence per 100000 per year
log(lambda[i]) <- inprod(alpha[],X[i,])
X[i,1] <- 1
for(k in 2:6){X[i,k] <- equals(age[i],k) }
for(k in 2:4){X[i,k+5] <- equals(city[i],k) }
}
for(k in 1:9){alpha[k] ~  dnorm(0,0.001);A[k] <- exp(alpha[k])}
}
```

```
# inits:
list(alpha=c(0,0,0,0,0,0,0,0,0))
```

4. Analyze the following data using gamma$(\mu\theta, \theta)$ density for the observations. Plot the patient specific means in comparison with the observed values using 'Comparison Tool'. Plot the means in comparison with WBC $\in [10^3, 10^5]$ for AG=1 and AG=2.

$$\log(\mu_i) = \alpha_0 + \alpha_1 I_{\{\mathrm{AG}_i=+\}} + \alpha_2 \log(\mathrm{WBC}_i)$$

WBC: The white blood cell count. Time: The time to death (in weeks). AG: AG is 1 for AG-positive patients, and is 2 for AG-negative patients.
`http://www.sci.usq.edu.au/staff/dunn/Datasets/Books/Hand/Hand-R/leukwbc-R.html`
(Also in R, type: `library(MASS), data(leuk), help(leuk)`)

WBC[]	Time[]	AG[]
2300	65	1
750	156	1
4300	100	1
2600	134	1
6000	16	1
10500	108	1
10000	121	1
17000	4	1
5400	39	1
7000	143	1
9400	56	1
32000	26	1
35000	22	1
100000	1	1
100000	1	1
52000	5	1
100000	65	1
4400	56	2
3000	65	2
4000	17	2
1500	7	2
9000	16	2
5300	22	2
10000	3	2
19000	4	2
27000	2	2
28000	3	2
31000	8	2
26000	4	2
21000	3	2

```
79000   30     2
100000  4      2
100000  43     2
END


model{
for(i in 1:33){
patient[i] <- i
Time[i]  ~ dgamma(a[i],theta)
a[i] <- mu[i]*theta
log(mu[i]) <- alpha[1]+alpha[2]*equals(AG[i],1)+alpha[3]*log(WBC[i])
var[i] <- mu[i] / theta
}
VMR <- 1 / theta   # variance to mean ratio
theta ~ dgamma(0.01,0.01)
for(u in 1:3){  alpha[u] ~ dnorm(0,0.001)    }
for(i in 1:100){
WB[i] <- i*1000; LWB[i] <- log(WB[i])
log(predmu[i,1]) <- alpha[1] + alpha[2] + alpha[3]*log(WB[i])
log(predmu[i,2]) <- alpha[1] + alpha[3]*log(WB[i])
}}
# inits:
list(alpha=c(0,0,0),theta=1)
```