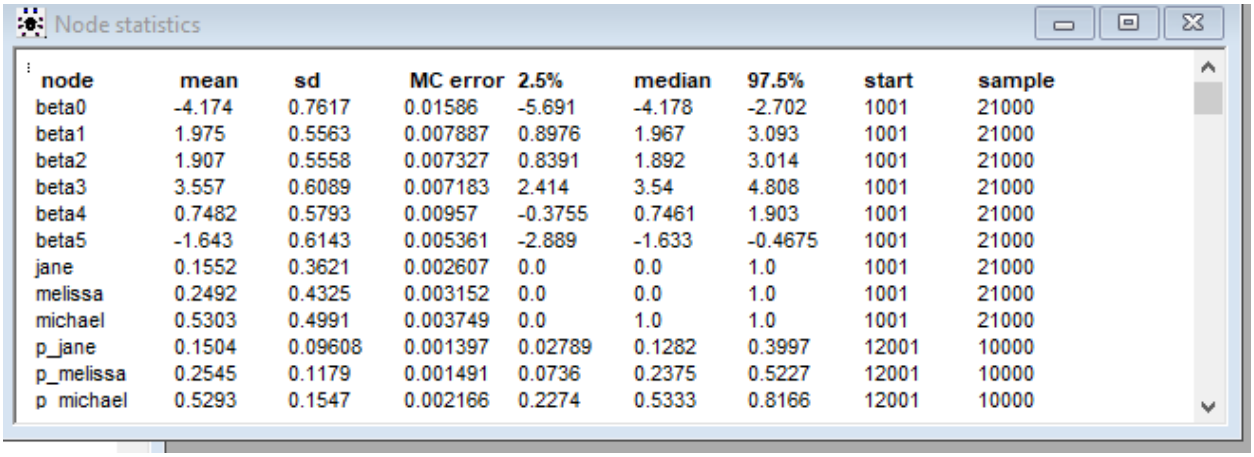
Kien Duc Vu Final Exam

Problem 1:

-95% credible set of probability of Jane is [0.02789, 0.3997] with mean = 0.1504. The probability of Janes going to the beach calculated in homework 1 is 0.17238, which is fell in 95% credible set.

-95% credible set of probability of Michael is [0.2274, 0.8166] which mean = 0.5293. The probability of Michael going to the beach calculated in homework 1 is 0.40704, which is fell into 95% credible set.

-95% credible set of probability of Melissa is [0.0736, 0.5227] which mean = 0.2545. The probability of Melissa going to the beach calculated in homework 1 is 0.27964 which is fell into 95% credible set.



node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	-4.174	0.7617	0.01586	-5.691	-4.178	-2.702	1001	21000
beta1	1.975	0.5563	0.007887	0.8976	1.967	3.093	1001	21000
beta2	1.907	0.5558	0.007327	0.8391	1.892	3.014	1001	21000
beta3	3.557	0.6089	0.007183	2.414	3.54	4.808	1001	21000
beta4	0.7482	0.5793	0.00957	-0.3755	0.7461	1.903	1001	21000
beta5	-1.643	0.6143	0.005361	-2.889	-1.633	-0.4675	1001	21000
jane	0.1552	0.3621	0.002607	0.0	0.0	1.0	1001	21000
melissa	0.2492	0.4325	0.003152	0.0	0.0	1.0	1001	21000
michael	0.5303	0.4991	0.003749	0.0	1.0	1.0	1001	21000
p_jane	0.1504	0.09608	0.001397	0.02789	0.1282	0.3997	12001	10000
p_melissa	0.2545	0.1179	0.001491	0.0736	0.2375	0.5227	12001	10000
p_michael	0.5293	0.1547	0.002166	0.2274	0.5333	0.8166	12001	10000

Problem 2:

a) 95 % credible set of LC50 predicted from provided values is [1.877, 6.617] with mean = 4.234. And the 95% credible set of mean of LC50 is [3.88, 4.591] which bar(miu) is 4.24.

b) H-050 and C-040 can be ignored while fitting Bayesian multilinear regression since their 95% credible set contain 0 and their means ~ 0.

H-050 95% credible set = [-0.07797, 0.1547] with mean = 0.03925

C-040 95% credible set = [-0.1517, 0.1551] with mean = 0.002343

c) Bayesian R-square = 0.4841 with 95% credible set [0.4192, 0.5426]. Since Bayesian R^2 is centered around 0.4841, the provided features to predict LC50 can only explain 50% variance in the prediction. Some of the predictors must be missing.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
BR2	0.4841	0.03132	3.108E-4	0.4192	0.4851	0.5426	1001	10000
BR2adj	0.4764	0.03179	3.155E-4	0.4106	0.4774	0.5358	1001	10000
beta0	2.702	0.2452	0.002691	2.217	2.701	3.179	1001	10000
beta1	0.02717	0.002657	2.568E-5	0.02203	0.02715	0.03249	1001	10000
beta2	-0.01505	0.0021	2.053E-5	-0.01924	-0.01507	-0.01093	1001	10000
beta3	0.03925	0.0596	5.831E-4	-0.07797	0.0394	0.1547	1001	10000
beta4	0.445	0.06335	5.884E-4	0.3224	0.4446	0.5698	1001	10000
beta5	0.5153	0.1353	0.001139	0.2444	0.5157	0.7754	1001	10000
beta6	-0.5738	0.1521	0.001555	-0.8785	-0.5747	-0.2732	1001	10000
beta7	-0.2248	0.04899	5.197E-4	-0.3206	-0.2252	-0.1294	1001	10000
beta8	0.002343	0.07846	8.254E-4	-0.1517	0.002848	0.1551	1001	10000
lc_pred	4.234	1.21	0.01267	1.877	4.236	6.617	1001	10000
miu	4.24	0.1812	0.002009	3.88	4.239	4.591	1001	10000

Problem 3:

a. We might believe that it is unreasonable to assume that all the studies in our meta-analysis are estimating exactly the same treatment effect, and that are they are the only studies in which we are interested, or perhaps the only studies that exist on the topic. We may assume that differences in samples, design, and conduct introduce more *statistical heterogeneity* between studies, than can reasonably be attributed to random error within studies. In that case it might be more reasonable to use a random effects model. A random effects model assumes the studies are a sample from all possible studies and includes an additional variance component for variation or heterogeneity between studies.

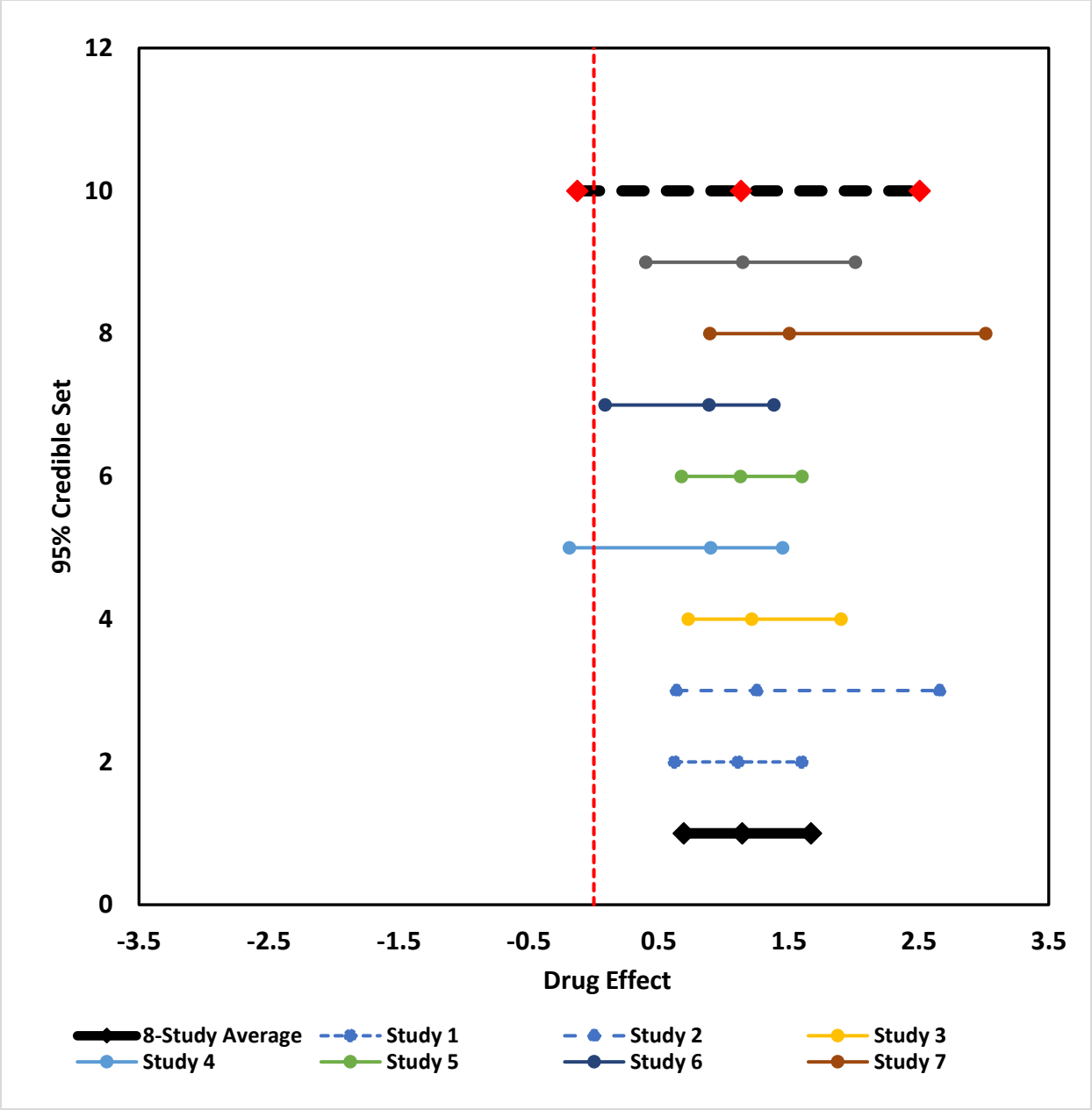
In this case, it is binomial distribution, where: (please see appendix for model)

- pc[i], pt[i] are the total number of patients in the two arms of the study, or the drug/placebo populations
- rc[i], rt[i] are the number of events in the two arms
- pc[i], pt[i] are the underlying probabilities used to define the likelihood. This is used to model in-between heterogeneity
- mu[i] is the log odds event in group A (and requires a prior)
- delta[i] is the so-called "treatment effect" or the log OR for the ith study, and
- d is the study-level point estimate, with tau the between study variance
- delta.new is the predictive probability distribution for most likely outcome in a future study

Node statistics

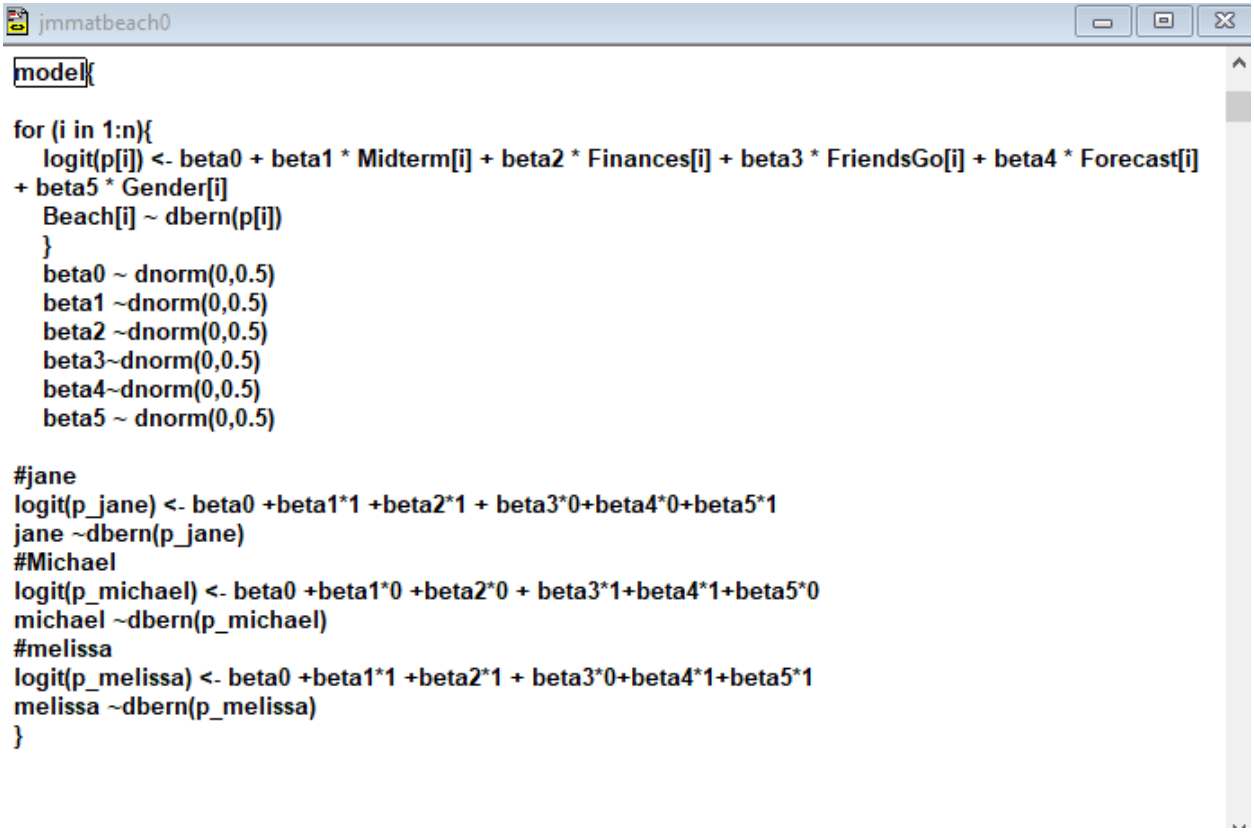
node	mean	sd	MC error	2.5%	median	97.5%	start	sample
OR.new	-40.7	5267.0	23.5	-12.28	-3.103	-0.8804	1001	50000
d	1.148	0.2438	0.003311	0.6905	1.14	1.671	1001	50000
delta[1]	1.108	0.2457	0.002991	0.6208	1.108	1.599	1001	50000
delta[2]	1.356	0.5025	0.00829	0.6347	1.254	2.661	1001	50000
delta[3]	1.241	0.2945	0.004278	0.7263	1.215	1.902	1001	50000
delta[4]	0.8137	0.423	0.006879	-0.1874	0.8991	1.452	1001	50000
delta[5]	1.13	0.2325	0.002378	0.6738	1.128	1.601	1001	50000
delta[6]	0.8408	0.3373	0.005399	0.08708	0.8869	1.386	1001	50000
delta[7]	1.64	0.567	0.01322	0.8921	1.505	3.016	1001	50000
delta[8]	1.161	0.3834	0.003833	0.3997	1.146	2.013	1001	50000
delta.new	1.15	0.6757	0.004647	-0.1269	1.132	2.508	1001	50000
sigma	0.3407	0.2828	0.006141	0.03239	0.2722	1.062	1001	50000
tau	108.7	314.5	5.834	0.8865	13.49	955.1	1001	50000

b.



Based on Bayesian meta-analysis accounting for in-between heterogeneity, drug treatment from 8 studies is most likely beneficial with 95% credible drug effect is [0.6905, 1.671] with mean of 1.14. Future study predictive drug treatment effect is most likely beneficial also, which 95% credible set of predictive future effect is [-0.1269, 2.508] with mean of 1.132.

Problem 1: Model



Problem 2 Model:

```
model{
 for (i in 1:n){
lc50[i] ~dnorm(mu[i], tau)
mu[i] <- beta0 + beta1*tpsa[i] + beta2*saacc[i] + beta3*h050[i] + beta4 *mlogp[i] + beta5*rdchi[i]
+ beta6*gats1p[i] + beta7*nn[i] +beta8*c040[i]
}
tau ~ dgamma(0.01,0.01)
sigma2 <-1/tau
beta0~ dnorm(0,0.001)
beta1~ dnorm(0,0.001)
beta2~ dnorm(0,0.001)
beta3~ dnorm(0,0.001)
beta4~ dnorm(0,0.001)
beta5~ dnorm(0,0.001)
beta6~ dnorm(0,0.001)
beta7~ dnorm(0,0.001)
beta8~ dnorm(0,0.001)
p <- 9

#prediction
lc_pred ~dnorm(miu,tau)
miu <- beta0 +beta1*12.5 + beta3*0.4 + beta4*1.5 + beta5*1 +beta2*0 + beta6*0 + beta7*0
+beta8*0

#Bayesian R^2
sse <- (n-p)*sigma2
for( i in 1:n){
cy[i] <- lc50[i] - mean(lc50[])}
sst <- inprod(cy[], cy[])
BR2 <- 1 - sse/sst
BR2adj <- 1- (n-1)*sigma2/sst
}
}
```

Problem 3-Model

metaanalysis

```
model
{
   for( i in 1 : Num ) {
      rc[i] ~ dbin(pc[i], nc[i])
      rt[i] ~ dbin(pt[i], nt[i])
      logit(pc[i]) <- mu[i] +delta[i]
      logit(pt[i]) <- mu[i]
      mu[i] ~ dnorm(0.0,1.0E-5)
      delta[i] ~ dt(d, tau, 4)
   }
   d ~ dnorm(0.0,1.0E-6)
   tau ~ dgamma(0.001,0.001)
   delta.new ~ dt(d, tau, 4)
  OR.new <- -exp(delta.new)
   sigma <- 1 / sqrt(tau)
}

list(rt = c(16, 1, 8, 9, 32, 15, 2, 3),
     nt = c(141, 53, 136, 59, 95, 107, 113,317),
     rc = c(41, 8, 28, 9, 59, 20, 27, 5),
     nc = c(152, 52, 139, 51, 97, 99, 132, 159),
     Num = 8)

list(d = 0, delta.new = 0, tau=1, mu = c(0, 0, 0, 0, 0, 0,
0, 0),
     delta = c(0, 0, 0, 0, 0, 0, 0, 0))
```