2. (a) The Poisson model is reasonable because the number of bikes represents finite count data and each hub can be considered an enclosed space.

Let $\underset{\sim}{y} \sim$ POISSON(θ),

Then θ represents the true average number of bikes per hub and each y_i represents the count of bikes at the ith hub, which each are assumed to come from the same distribution, with the same mean (θ).

(b) The Gamma distribution is often chosen as a prior on a POISSON parameter. The POISSON parameter represents the mean as well as the variation in the bike counts. Since variances are always positive, and the family of Gamma distributions have a support of (0,infinite), it is a reasonable choice. While choosing a member of the Gamma distributions on θ is convenient because it serves as a conjugate prior, that is not the justification I am using. It is, however, a nice property.

α can be thought of as the prior count and β can be thought of as the prior number of observations. I do not have any prior knowledge, so I will set the prior number of observations to 1, and make the distribution improper by choosing the prior number of counts as 0. Doing so makes the prior distribution flat, as also can be seen in the first two moments.

(c) $p(\theta|\underset{\sim}{y}) \sim \text{GAMMA}(\Sigma y_i + \alpha, n + \beta)$

when $y_i \sim$ POISSON(θ) and

$\theta \sim \text{GAMMA}(\alpha, \beta)$.

p$(\theta|\underset{\sim}{y}) \propto \int_\theta p(\underset{\sim}{y}|\theta) \times p(\theta)$ dθ

$\propto \int_\theta \frac{\theta^{\Sigma y_i} e^{-n\theta}}{\Pi y_i!} \times \frac{\beta^\alpha \theta^{\alpha-1} e^{-\theta\beta}}{\Gamma(\alpha)} d\theta$

$\propto \int_\theta \frac{\theta^{\Sigma y_i + \alpha - 1} e^{-\theta(n+\beta)} \beta^\alpha}{\Pi y_i \Gamma(\alpha)} d\theta$

$\propto$ GAMMA($\Sigma y_i + \alpha$, $n + \beta$)

$\propto$ GAMMA(9435 + 0, 392 + 1)

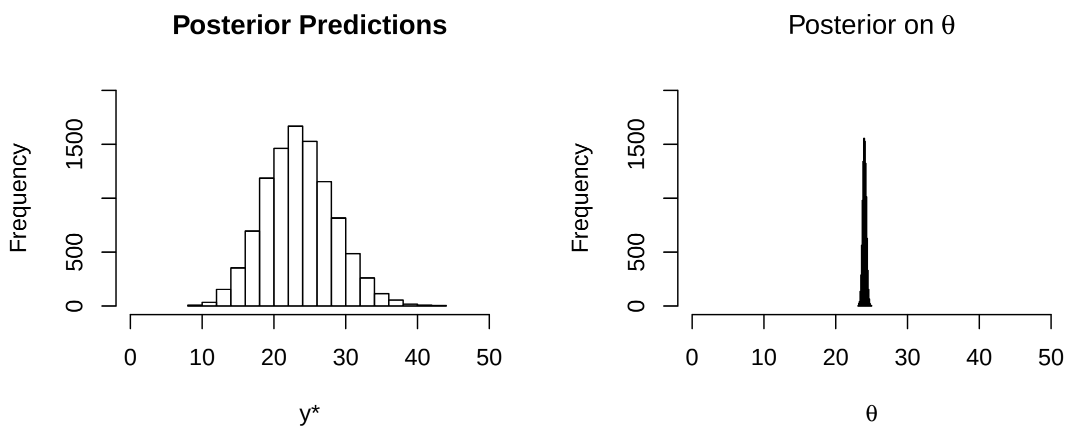
$\propto$ GAMMA(9435, 393)

(d) p$(y^*|\underset{\sim}{y}) = \int_\theta p(\Sigma y^*|\theta) \times p(\theta|\underset{\sim}{y}) d\theta$

From the notes, p$(y^*|\underset{\sim}{y}) \sim \text{NEGBINOM}(\Sigma y_i + \alpha, n + \beta)$

But rather than drawing samples from the Negative Binomial model, we can use the MC algorithm to draw predictions using random draws from the posterior of θ and the conditional, $y^*|\theta \sim$ POISSON(θ).

The posterior predicted values generated are plotted below. There is much less variation in the posterior of θ than in the posterior predictions. Since we are using two draws for each prediction (one from the posterior on θ and one from the predictive on y^*), it makes sense that there should be more variation in the y^* predictive distribution. Both have a fairly symmetrical shape and nearly the same center. I am surprised with the degree of accuracy shown in the posterior of θ compared to the predictions of y^*.



(e) The 95% credible interval is estimated to be (2.9407029, 3.061808) bikes per slot and the point estimate is 3.00102 bikes per slot.

(f) The MAP point estimate is within the interval (24.18,24.19] while the mean of the posterior distribution should theoretically be 24.0076336 and the mean based on the simulations was 24.0081603. Bother are less than the MAP point estimate. No the mean is not the MAP point estimate.

(g) The estimated highest 95% posterior density region is (23.33, 24.65] compared to the 95% credible interval of (23.5256232, 24.4944641). My bins were 0.01 units in estimating the HPD region.

My HPD region is slightly wider than the credible interval and was estimated as the 94.96% HPD region. It is likely the differences are due to the size of my bins so yes, I would say the 95% HPD region is the 95% credible interval here.

3. (a) I do not have any prior information. So I will choose a weakly informative prior, and use the same one for both sets of years, as I do not have any more information about one year over the other.

The prior mean for both sets of years will be set to 0 and the prior variance for both sets of years will be set to 100. The prior variance is set to this based on the empirical rule, we expect about 95% of the data to lie within 2sd's of the mean. I'll say I expect to be within about 20 dollars of the true mean.

I will generalize with θ and σ^2 since I am assuming the same prior for both time periods.

$\theta \sim$ N(0, 100)

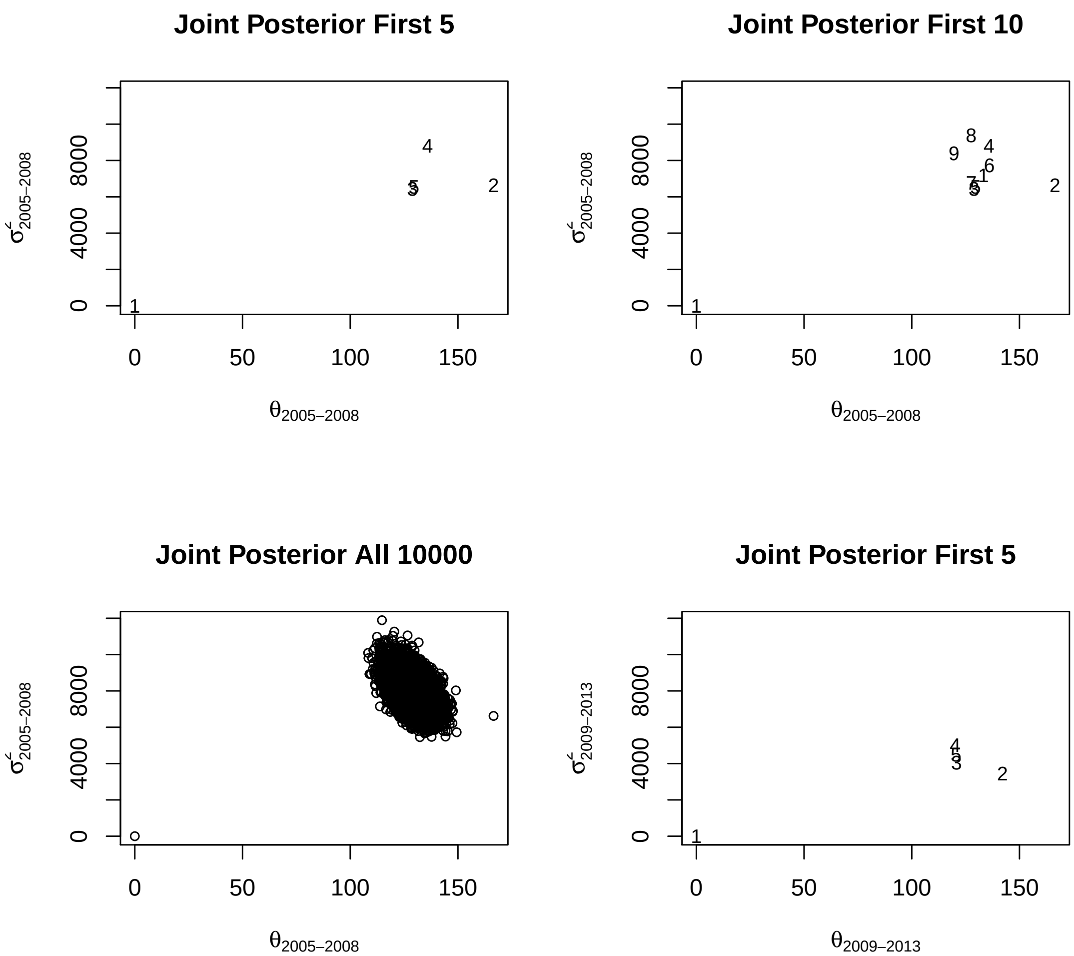
Where ν_o represents the prior number of observations and σ_o^2 represents the prior sample variance, I will set these to 1 and 100 respectively, using 100 with the same idea as above in mind.

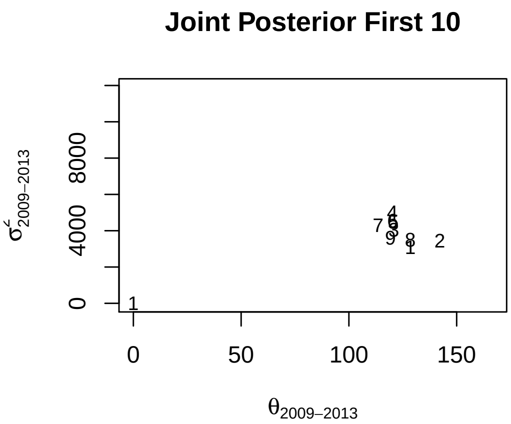
$\sigma^2 \sim$ INVGAM(0.5,50)

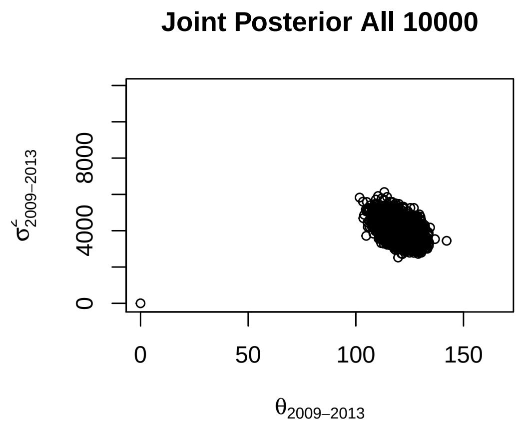
$\implies \frac{1}{\sigma^2} \sim$ INVGAM(0.5,50)

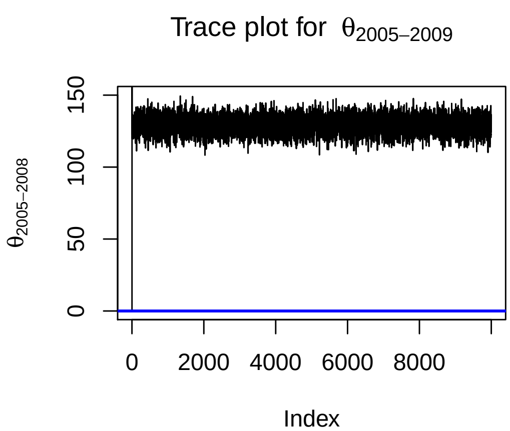
(b) Below are plots to assess the covergence of the joint posterior as well as the marginals for both sets of years.

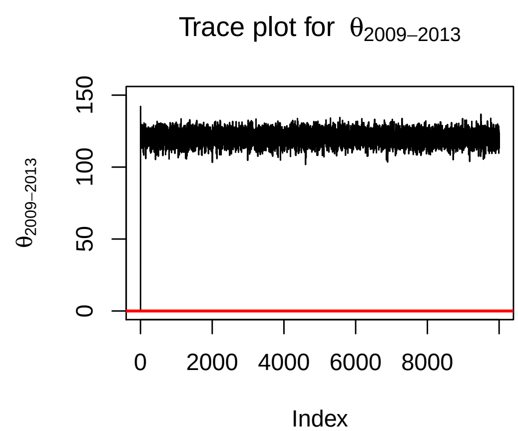
Both joint posteriors appear to converge within the first few iterations.

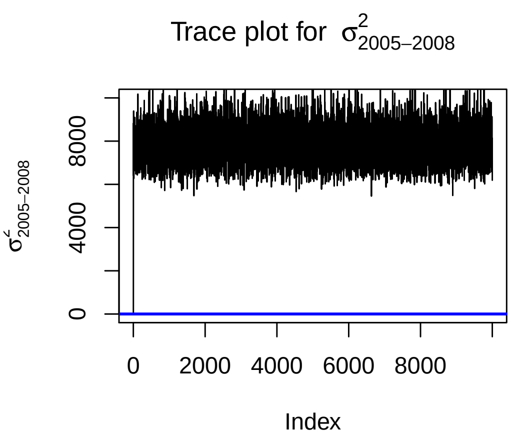


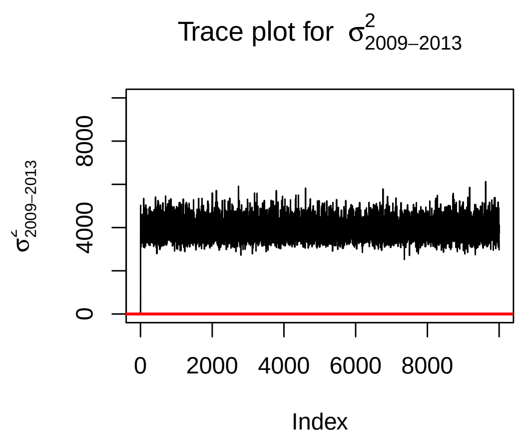

Joint Posterior First 10
σ²2009–2013
θ2009–2013
0
4000
8000
0
50
100
150


Joint Posterior All 10000
σ²2009–2013
θ2009–2013
0
4000
8000
0
50
100
150

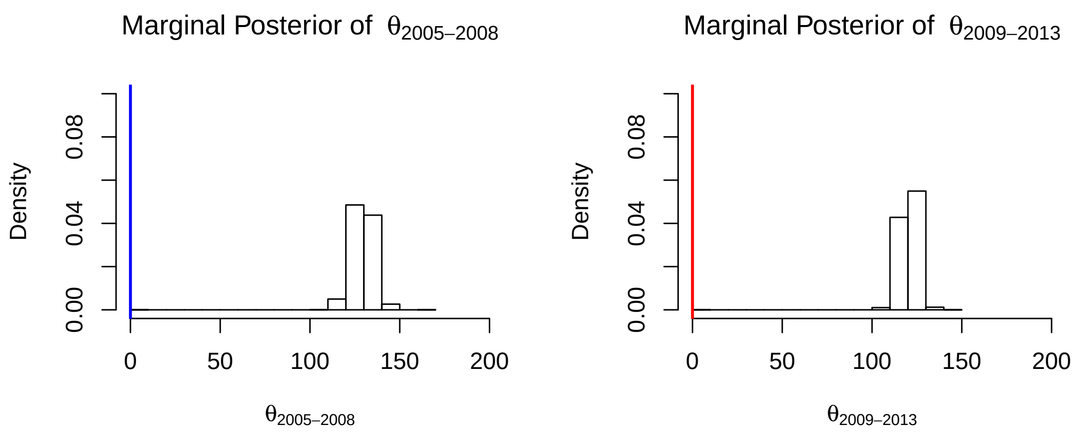

Trace plot for θ2005–2009
θ2005–2008
Index
0
50
100
150
0
2000
4000
6000
8000


Trace plot for θ2009–2013
θ2009–2013
Index
0
50
100
150
0
2000
4000
6000
8000


Trace plot for σ²2005–2008
σ²2005–2008
Index
0
4000
8000
0
2000
4000
6000
8000


Trace plot for σ²2009–2013
σ²2009–2013
Index
0
4000
8000
0
2000
4000
6000
8000

(c) Below are the marginal posteriors of the θs.



(d) Where c = $\$10/ft^2$, the $\Pr(|\theta_2 - \theta_1| > c)$ is estimated to be 0.2196 using the GIBBs algorithm.

(e) In part (d) I found that 21.96% of the iterations were more than $\$10/ft^2$ apart. Since this is a Bayesian framework, we would say there is a 21.96% that a change point has occured.

The observed difference in simulated posterior means was 8.7649131 with added individual variances of 54.0511367, making the standard deviation around 7. Note, we can only do this if we assume the two time periods are independent, which they surely are not, and a covariance term could increase or decrease this and with the variable housing markets, I'm not sure which would be more accurate. Considering the size of the observed difference, there is quite a bit of uncertainty. But I think relative to all of this information, the interpretation of the probability of a point change is representative of the uncertainty and the difference in means.

R Code

```
require(xtable)
require(ggplot2)
require(dplyr) #between()
require(truncnorm) #work with truncated norm
require(LearnBayes)
opts_chunk$set(echo = FALSE, warning = FALSE, fig.align = 'center', fig.height = 3.5, fig.width = 4)
```

```
bikes <- read.csv("MidtermBikes.csv")
head(bikes)
```

3. (a)

```
yi <- bikes$bikes.rented
sum.yi <- sum(yi)
n.yi <- length(yi)

alpha = 0
beta = 1
```

```
#posterior on theta
set.seed(532)
theta_i <- rgamma(10000, alpha + sum.yi, beta + n.yi)
```

(b)

```
#posterior predictive
ystar <- c(rep(0,10000))

for(i in 1:10000){
ystar[i] <- rpois(1, theta_i[i])
}

par(mfrow=c(1,2))
hist(ystar, main = "Posterior Predictions", xlab =
       "y*", xlim=c(0,50), ylim = c(0,2000))
hist(theta_i, main = expression("Posterior on" ~ theta),
```

```
    xlab = expression(theta), xlim=c(0,50), ylim = c(0,2000))
```

(c)
```
lb <- qgamma(0.025, shape = alpha + sum.yi, rate = beta + n.yi)

ub <- qgamma(0.975, shape = alpha + sum.yi, rate = beta + n.yi)

point.est <- mean(theta_i)/8
```

(d)
```
theta_i.binned <- cut(theta_i, breaks = c(seq(range(theta_i)[1]-0.01,range(theta_i)[2]+0.01, by = 0.01)))
map.theta_i <- theta_i.binned[max(tabulate(theta_i.binned))]
mean.theta_i.sim <- mean(theta_i)
mean.theta_i.theoretical <- (alpha + sum.yi)/(n.yi+beta)
```

(e)
```
it <- table(theta_i.binned)

it1 <- it - 1

pd1 <- (sum(it1[it1>=0]) - sum(it1))/length(theta_i)

it2 <- it - 2

pd2 <- (sum(it2[it2>=0]) - sum(it2))/length(theta_i)

it3 <- it - 3

pd3 <- (sum(it3[it3>=0]) - sum(it3))/length(theta_i)

it5 <- it - 5

pd5 <- (sum(it5[it5>=0]) - sum(it5))/length(theta_i)

it8 <- it - 8

pd8 <- (sum(it8[it8>=0]) - sum(it8))/length(theta_i)

it7 <- it - 7

pd7 <- (sum(it7[it7>=0]) - sum(it7))/length(theta_i)

it6 <- it - 6

pd6 <- (sum(it6[it6>=0]) - sum(it6))/length(theta_i)

#pd7 is very close to a 95% hpd

lb.hpd <- it[which(it>=7)][1]#23.33
ub.hpd <- it[which(it>=7)][length(which(it>=7))]#24.65
```

4. (a)
```
bzh <- read.csv("BozemanHousing.csv")
head(bzh)

bzh0508 <- subset(bzh, bzh$YearSold == "2005-2008")
bzh0913 <- subset(bzh, bzh$YearSold == "2009-2013")

s2 <- 100
prior.mean <- 0
```

(b)
```
num.sims <- 10000
Phi0508 <- matrix(0,nrow=num.sims,ncol=2)
Phi0508[1,1] <- 0 # initialize theta
Phi0508[1,2] <- 1 # initialize (1/sigmasq)
mu.0 <- 0
tausq.0 <- 100
nu.0 <- 1
sigmasq.0 <- 100
num.obs0508 <- dim(bzh0508)[1]

mean.y0508 <- mean(bzh0508$Closing_Price_per_sqft)

for (i in 2:num.sims){
  # sample theta from full conditional
  mu.n0508 <- (mu.0 / tausq.0 + num.obs0508 * mean.y0508 *Phi0508[(i-1),2]) / (1 / tausq.0 + num.obs0508 * Phi0508[(i-1),2] )
  tausq.n0508 <- 1 / (1/tausq.0 + num.obs0508 * Phi0508[(i-1),2])
  Phi0508[i,1] <- rnorm(1,mu.n0508,sqrt(tausq.n0508))

  # sample (1/sigma.sq) from full conditional
  nu.n0508 <- nu.0 + num.obs0508
  sigmasq.n.theta0508 <- 1/nu.n0508*(nu.0*sigmasq.0 + sum((bzh0508$Closing_Price_per_sqft - Phi0508[i,1])^2))
  Phi0508[i,2] <- rgamma(1,nu.n0508/2,nu.n0508*sigmasq.n.theta0508/2)
}

#next 2009-2013

num.sims <- 10000
Phi0913 <- matrix(0,nrow=num.sims,ncol=2)
Phi0913[1,1] <- 0 # initialize theta
Phi0913[1,2] <- 1 # initialize (1/sigmasq)
mu.0 <- 0
tausq.0 <- 100
nu.0 <- 1
sigmasq.0 <- 100
num.obs0913 <- dim(bzh0913)[1]

mean.y0913 <- mean(bzh0913$Closing_Price_per_sqft)

for (i in 2:num.sims){
  # sample theta from full conditional
  mu.n0913 <- (mu.0 / tausq.0 + num.obs0913 * mean.y0913 *Phi0913[(i-1),2]) / (1 / tausq.0 + num.obs0913 * Phi0913[(i-1),2] )
```

```
  tausq.n0913 <- 1 / (1/tausq.0 + num.obs0913 * Phi0913[(i-1),2])
  Phi0913[i,1] <- rnorm(1,mu.n0913,sqrt(tausq.n0913))

  # sample (1/sigma.sq) from full conditional
  nu.n0913 <- nu.0 + num.obs0913
  sigmasq.n.theta0913 <- 1/nu.n0913*(nu.0*sigmasq.0 + sum((bzh0913$Closing_Price_per_sqft - Phi0913[i,1])^2))
  Phi0913[i,2] <- rgamma(1,nu.n0913/2,nu.n0913*sigmasq.n.theta0913/2)
}
par(mfrow=c(1,2))

# plot joint posterior 2005-2008
plot(Phi0508[1:5,1],1/Phi0508[1:5,2],xlim=range(Phi0508[,1]),ylim=range(1/Phi0508[,2]),cex=.8,pch=as.character(1:5), ylab=expression(sigma [2005-2008]^2), xlab = expression(theta [2005-20

plot(Phi0508[1:10,1],1/Phi0508[1:10,2],xlim=range(Phi0508[,1]),ylim=range(1/Phi0508[,2]),cex=.8,pch=as.character(1:10), ylab=expression(sigma [2005-2008]^2), xlab = expression(theta [2005

plot(Phi0508[,1],1/Phi0508[,2],xlim=range(Phi0508[,1]),ylim=range(1/Phi0508[,2]),cex=.8, ylab=expression(sigma [2005-2008]^2), xlab = expression(theta [2005-2008]), main='Joint Posterior


# plot joint posterior 2009-2013
plot(Phi0913[1:5,1],1/Phi0913[1:5,2],xlim=range(Phi0508[,1]),ylim=range(1/Phi0508[,2]),cex=.8,pch=as.character(1:5), ylab=expression(sigma [2009-2013]^2), xlab = expression(theta [2009-20

plot(Phi0913[1:10,1],1/Phi0913[1:10,2],xlim=range(Phi0508[,1]),ylim=range(1/Phi0508[,2]),cex=.8,pch=as.character(1:10), ylab=expression(sigma [2009-2013]^2), xlab = expression(theta [2009

plot(Phi0913[,1],1/Phi0913[,2],xlim=range(Phi0508[,1]),ylim=range(1/Phi0508[,2]),cex=.8, ylab=expression(sigma [2009-2013]^2), xlab = expression(theta [2009-2013]), main='Joint Posterior


# plot trace plots theta 2005-2008
plot(Phi0508[,1],type='l',ylab=expression(theta [2005-2008]), main=expression('Trace plot for ' ~ theta [2005-2009]), ylim = c(0,150))
abline(h=0,lwd=2,col='blue')

# plot trace plots theta 2009-2013
plot(Phi0913[,1],type='l',ylab=expression(theta [2009-2013]), main=expression('Trace plot for ' ~ theta [2009-2013]), ylim = c(0,150))
abline(h=0,lwd=2,col='red')

# plot trace plots s2 2005-2008
plot(1/Phi0508[,2],type='l',ylab=expression(sigma[2005-2008]^2), main=expression('Trace plot for ' ~ sigma[2005-2008]^2), ylim = c(0,10000))
abline(h=1,lwd=2,col='blue')


# plot trace plots s2 2009-2013
plot(1/Phi0913[,2],type='l',ylab=expression(sigma[2009-2013]^2), main=expression('Trace plot for ' ~ sigma[2009-2013]^2), ylim = c(0,10000))
abline(h=1,lwd=2,col='red')
```

(c)

```
par(mfrow=c(1,2))
# plot marginal posterior of theta 2005-2008
hist(Phi0508[,1],xlab=expression(theta [2005-2008]),main=expression('Marginal Posterior of ' ~ theta [2005-2008]),probability=T, ylim = c(0,0.1), xlim = c(0,200))
abline(v=0,col='blue',lwd=2)

# plot marginal posterior of theta 2009-2013
hist(Phi0913[,1],xlab=expression(theta [2009-2013]),main=expression('Marginal Posterior of ' ~ theta [2009-2013]),probability=T, ylim = c(0,0.1), xlim = c(0,200))
abline(v=0,col='red',lwd=2)

# plot marginal posterior of sigmasq 2005-2008
#hist(1/Phi0508[,2],xlab=expression(sigma[2005-2008]^2),main=expression('Marginal Posterior of ' ~ sigma[2005-2008]^2),probability=T, ylim = c(0,0.001), xlim = c(0,10000))
#abline(v=1,col='blue',lwd=2)

# plot marginal posterior of sigmasq 2009-2013
#hist(1/Phi0913[,2],xlab=expression(sigma[2009-2013]^2),main=expression('Marginal Posterior of ' ~ sigma[2009-2013]^2),probability=T, ylim = c(0,0.001), xlim = c(0,10000))
#abline(v=1,col='red',lwd=2)
```

(d)

```
c.ft <- 10
prob.est <- mean(c(mean(Phi0508[,1] > Phi0913[,1] + c.ft), mean(Phi0508[,1] < Phi0913[,1] - c.ft)))
```