**ReadMe for the R – Code**

**Importing the data into R**

#Bayesian Statistics Project

#Drug Overdose Deaths Analysis

#import data into R

setwd("/Users/victoragboli/Documents/Spring 2023 Semester/Bayesian/My Bayesian/Project")

df = read.csv("Drug Overdose Deaths.csv")

**Check data type of each column**

#Check data type

str(df)

**Removing the data frame headers**

colnames(df) = NULL#Removing the data frame headers

head(df, 10) #inspecting the first 10 rows

**Beginning of 2017 vs. 2018 Analysis**

#2017 vs. 2018 Analysis

pre = df[, 5]

post = df[, 4]

**Viewing the histogram**

#histogram

par(mfrow=c(2,2))

hist(pre,50,main="Drug Overdose Deaths in 2017",col="blue")

hist(post,50,main="Drug Overdose Deaths in 2018",col="red")

**Log transforming the 2017 and 2018 data**

#log transform

pre = log(pre)

post = log(post)

**Mean of 2017 and 2018 data**

mean(pre)

mean(post)

**Viewing the histogram of log transformation**

#histogram of log transformation

hist(pre,50,main="Drug Overdose Deaths in 2017",col="blue")

hist(post,50,main="Drug Overdose Deaths in 2018",col="red")

**The Likelihood Function**

#Likelihood Function

like=function(th){

mu1=th[1]; sig1=th[2]; mu2=th[3]; sig2=th[4]

prod(dnorm(pre, mean=mu1,sd=sig1))\*prod(dnorm(post,mean=mu2,sd=sig2))

}

**The Prior Distribution**

#prior Distribution

Prior=function(th){

mu1=th[1]; sig1=th[2]; mu2=th[3]; sig2=th[4]

if (sig1<=0 | sig2<=0) return(0)

dnorm(mu1,6.658,6.658)\*dnorm(mu2,6.622,6.622)\*dexp(sig1,rate=1/6.658)\*dexp(sig2,rate=1/6.622)

}

**The Posterior Distribution**

#posterior

Posterior=function(th){Prior(th)\*like(th)}

#starting

mu1=6.658; sig1=6.658; mu2=6.622; sig2=6.622

th0=c(mu1,sig1,mu2,sig2)

nit=1000000

results=matrix(0,nrow=nit,ncol=4)

th=th0

results[1,]=th0

for (it in 2:nit){

Cand=th + rnorm(4,sd=.003)

ratio=Posterior(Cand)/Posterior(th)

if (runif(1) < ratio) th=Cand

results[it,]=th

}

**Viewing the results**

#edit(results)

#getting the trace-plot

par(mfrow=c(4,1))

plot(results[,1])

plot(results[,2])

plot(results[,3])

plot(results[,4])

**Removing the burns**

#removing the burns from the traceplots

res=results[2.7e+05:1e+06,]

**Getting the traceplots**

par(mfrow=c(4,1))

plot(res[,1])

plot(res[,2])

plot(res[,3])

plot(res[,4])

mu1s=res[,1]

sig1s=res[,2]

mu2s=res[,3]

sig2s=res[,4]

par(mfrow=c(2,1))

plot(mu1s-mu2s)

**Getting the histogram**

hist(mu1s-mu2s)

**Getting the probability of mu1 < mu2**

mean(mu1s-mu2s<0)

**Beginning of 2019 vs. 2020 Analysis**

#2019 vs. 2020 Analysis

pre = df[, 3]

post = df[, 2]

**Viewing the histogram of 2019 and 2020**

#histogram

par(mfrow=c(2,2))

hist(pre,50,main="Drug Overdose Deaths in 2019",col="blue")

hist(post,50,main="Drug Overdose Deaths in 2020",col="red")

**Log transforming the 2019 and 2020 data**

#log transform

pre = log(pre)

post = log(post)

**Mean of 2019 and 2020 data**

mean(pre)

mean(post)

**Viewing the histogram of log transformation**

#histogram of log transformation

hist(pre,50,main="Drug Overdose Deaths in 2019",col="blue")

hist(post,50,main="Drug Overdose Deaths in 2020",col="red")

**The Likelihood Function**

#Likelihood Function

like=function(th){

mu1=th[1]; sig1=th[2]; mu2=th[3]; sig2=th[4]

prod(dnorm(pre, mean=mu1,sd=sig1))\*prod(dnorm(post,mean=mu2,sd=sig2))

}

**The Prior Distribution**

#prior Distribution

Prior=function(th){

mu1=th[1]; sig1=th[2]; mu2=th[3]; sig2=th[4]

if (sig1<=0 | sig2<=0) return(0)

dnorm(mu1,6.687,6.687)\*dnorm(mu2,6.929,6.929)\*dexp(sig1,rate=1/6.687)\*dexp(sig2,rate=1/6.929)

}

**The Posterior Distribution**

#posterior

Posterior=function(th){Prior(th)\*like(th)}

#starting

mu1=6.687; sig1=6.687; mu2=6.929; sig2=6.929

th0=c(mu1,sig1,mu2,sig2)

nit=1000000

results=matrix(0,nrow=nit,ncol=4)

th=th0

results[1,]=th0

for (it in 2:nit){

Cand=th + rnorm(4,sd=.003)

ratio=Posterior(Cand)/Posterior(th)

if (runif(1) < ratio) th=Cand

results[it,]=th

}

**Viewing the results**

#edit(results)

#getting the trace-plot

par(mfrow=c(4,1))

plot(results[,1])

plot(results[,2])

plot(results[,3])

plot(results[,4])

**Removing the burns**

#removing the burns from the traceplots

res=results[2.7e+05:1e+06,]

**Getting the traceplots**

par(mfrow=c(4,1))

plot(res[,1])

plot(res[,2])

plot(res[,3])

plot(res[,4])

mu1s=res[,1]

sig1s=res[,2]

mu2s=res[,3]

sig2s=res[,4]

par(mfrow=c(2,1))

plot(mu1s-mu2s)

**Getting the histogram**

hist(mu1s-mu2s)

**Getting the probability of mu1 < mu2**

mean(mu1s-mu2s<0)