Exam_20210603_Ali

Mohamed Ali - Mohal954

2023-05-24

Question 1

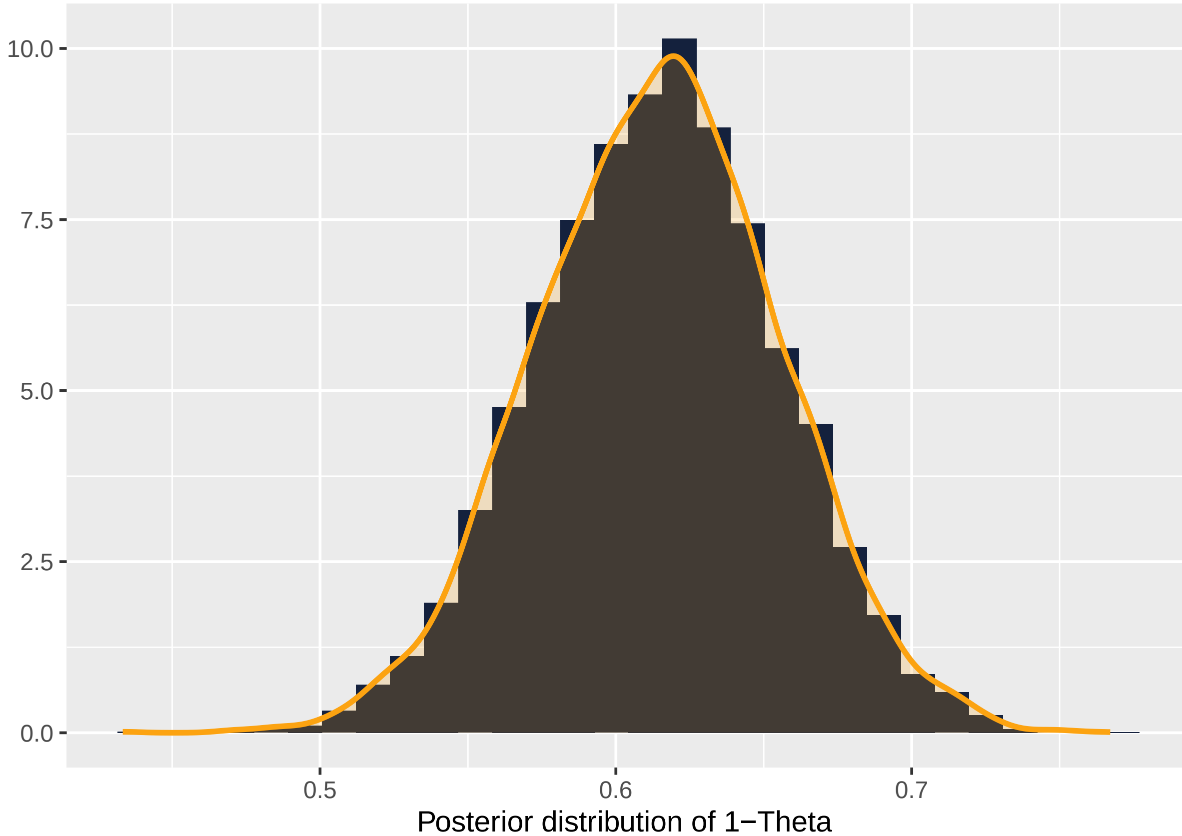
A

Since we amusing that we only have 2 chooses either the customer choose brand A or else then we are looking at a Bernoulli type of model distribution, where the measure the success by selecting brand A and fail by not selecting it The posterior in this case could be drives by applying the bayse role, P(theta|X) Prob-to P(X|theta)P(theta) the P(theta) here is beta(a,b), thus the posterior will follow beta(a+s,b+f). The derivations of the posterior could be found in my paper notes attached to this exam paper. We not define our S, and f.

```
n=100
A=38
B=27
C=35
alpha= 16
beta= 24

s=A
f=n-s


theta_A<- data.frame(x=rbeta(10000,alpha+s,beta+f))
theta_A$x2 <- 1-theta_A$x

ggplot(theta_A,aes(x = x2)) +geom_histogram(aes(y=..density..),
                                   linetype=1, fill='#14213D')+
  geom_density(alpha=.2,color="#FCA311",size=1,fill="#FCA311")+
  labs(x='Posterior distribution of 1-Theta ',y=' ',)
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.

## Warning: The dot-dot notation (‘..density..‘) was deprecated in ggplot2 3.4.0.
## i Please use ‘after_stat(density)‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.




```
paste0("Posterior probability of Theta >0.4 is : ",mean(theta_A>.4))
```

```
## [1] "Posterior probability of Theta >0.4 is : 0.6811"
```

B

```
theta_A$x3<- theta_A$x2/theta_A$x

mean(theta_A$x3)
```

```
## [1] 1.62228
```

```
quantile(theta_A$x3,c(0.05,0.95))
```

```
##       5%      95%
## 1.207031 2.137852
```

Form the above results we can see that the mean of the ratio 1-theta/theta is fall between the upper and the lower boundaries of the CI. which mean it's significant and can interpreted that 1.6 of the customer bought from a brand either than A . this could be sound if we're to look at just the ration of customers who selected brand not A = 62 and those who selected A = 38 and by taking he divition 62/38 = 1.63 which tell us this result is significant.

C

TO find the marginal distribution id to find the integral from 0 to 1 for theta^alpha-1 and theta^beta-1 d(theta) this could be found by using the function beta and we margnilize by clculating it for the alpha and beta old vs new

```
beta(alpha+s,beta+f)/beta(alpha,beta)
```

```
## [1] 7.556771e-30
```

D

Since the The Dirichlet distribution is the multivariate generalization of the univariate beta distribution. Its probability density function returns the belief that the probabilities of kk rival events are theta_j given that each event has been observed alpha_j-1 times. Here we gonna use the function rdirichlet from the package LaplacesDemon, which can give us the random deviates for theta_A, B and C.

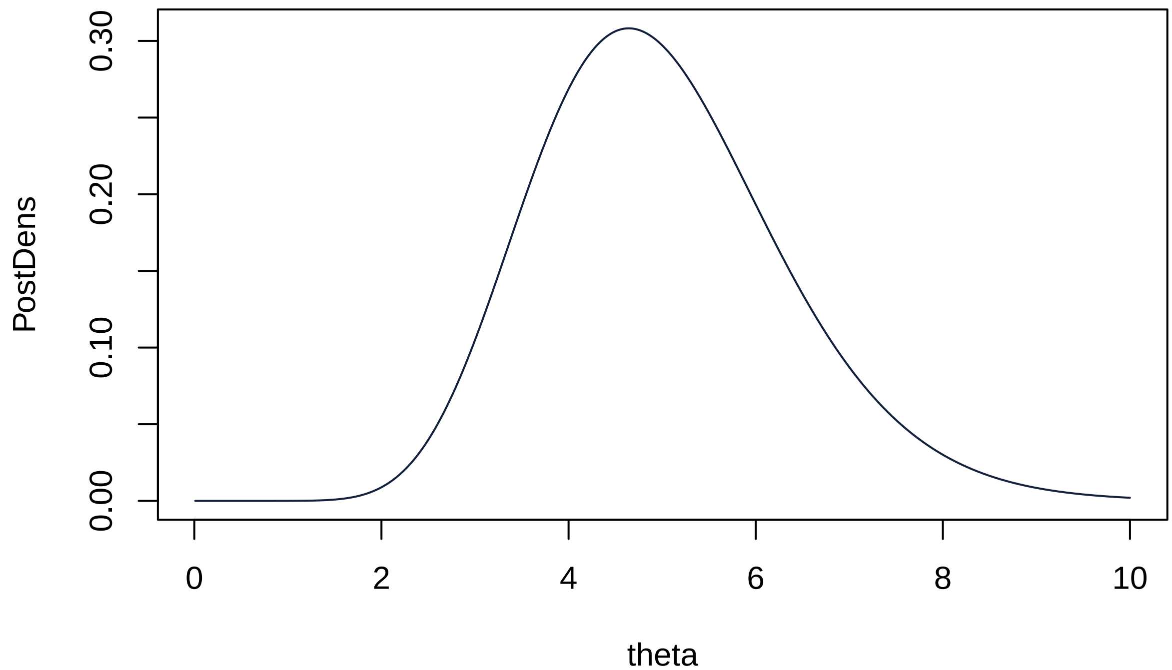
```
# first we generate a vector of all shape paramters, theta from the
# distribution beta with the alpha+s and beta+f then we use this vector to
# find the variates of theta_A,B and C

theta_draws<-rbeta(10000,alpha+s,beta+f)
rdir_res<-rdirichlet(3,theta_draws)
# the results is a matrix with three colun each column represent the theta for A, B and C respectivly.
mean(rdir_res[,1]>rdir_res[,2])
```

```
## [1] 0.6666667
```

Question 2

D

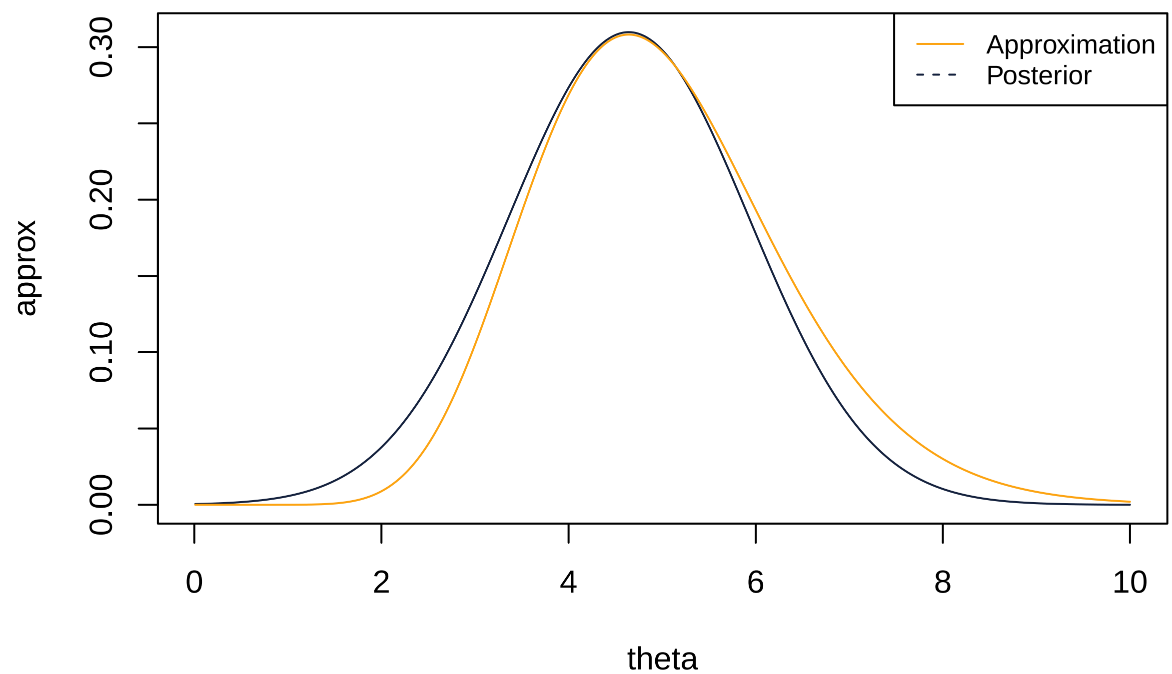
```
logpost<- function(theta,n,sum_xi2,lambda){
  loglikle<- n*log(theta) - (theta*sum_xi2)
  logprior<- log(lambda) - (lambda*sum_xi2)
  return(loglikle+logprior)
}

n=13
lambda=1/2
sum_xi2=2.8

theta <- seq(0.01,10,0.01)
PostDens_propto <- exp(logpost(theta,n,sum_xi2,lambda))
PostDens <- PostDens_propto/(0.01*sum(PostDens_propto))

plot(theta,PostDens,col="#14213D",type="l")
```



E

```
# Our logpost function
logpost<- function(theta,n,sum_xi2,lambda){
  loglikle<- n*log(theta) - (theta*sum_xi2)
  logprior<- log(lambda) - (lambda*sum_xi2)
  return(loglikle+logprior)
}

# initalize a seq of theta, always check the condistions of the theta
theta=seq(0.01,10,0.01)

# Our Pramaters for this function not used for other functions
n=13
lambda=1/2
sum_xi2=2.8

# Initial value for optim functoin
initVal <- 0

# we use optim function remember to set lower=.1 and method L-BFGS-B
optim_res<-optim(initVal,logpost,gr=NULL,n,sum_xi2,lambda,method=c("L-BFGS-B")
      ,control=list(fnscale=-1),hessian=TRUE,lower = 0.1)
```

```r
# mu extracted from the results of optim
mu<-optim_res$par
# Sigma values extracted from optim
sigma<- optim_res$hessian[1,1]
# we use dnorm to generate the probs from normal distribution
approx<-dnorm(theta,mean=mu,sd=sqrt(-1/sigma))

# results of the posterior from the preivous question

PostDens_propto <- exp(logpost(theta,n,sum_xi2,lambda))
PostDens <- PostDens_propto/(0.01*sum(PostDens_propto))

# Plotting part make sure to use this one instead of ggplot2
plot(theta,approx,col="#14213D",type="l")
lines(theta,PostDens,col="#FCA311")
legend("topright",legend = c("Approximation","Posterior"),
       col=c("#FCA311","#14213D"), lty=1:2, cex=0.8)
```



Question 3

A

```r
y=y
X=X
mu_0=rep(0,ncol(X))
Omega_0=(1/5^2)*diag(7)
v_0=1
sigma2_0=2^2
nIter=10000

betaSample<-BayesLinReg(y, X, mu_0, Omega_0, v_0, sigma2_0, nIter)$betaSample

for (i in 1:ncol(X)) {
  print(paste0("The Posterior Mean for beta_",i-1," is = "
         ,round(mean(betaSample[,i]),3)," and the 95% CI lower is= "
         ,round(quantile(betaSample[,i],c(0.05)),3)," and the 95% CI upper is= "
         ,round(quantile(betaSample[,i],c(.95)),3)))
}
```

```
## [1] "The Posterior Mean for beta_0 is = 1.308 and the 95% CI lower is= 1.177 and the 95% CI upper is=
## [1] "The Posterior Mean for beta_1 is = 0.7 and the 95% CI lower is= 0.555 and the 95% CI upper is= 0
## [1] "The Posterior Mean for beta_2 is = 0.157 and the 95% CI lower is= 0.066 and the 95% CI upper is=
## [1] "The Posterior Mean for beta_3 is = 0.425 and the 95% CI lower is= 0.089 and the 95% CI upper is=
## [1] "The Posterior Mean for beta_4 is = -0.162 and the 95% CI lower is= -0.336 and the 95% CI upper i
## [1] "The Posterior Mean for beta_5 is = 0.078 and the 95% CI lower is= -0.222 and the 95% CI upper is
## [1] "The Posterior Mean for beta_6 is = -0.239 and the 95% CI lower is= -0.416 and the 95% CI upper i
```

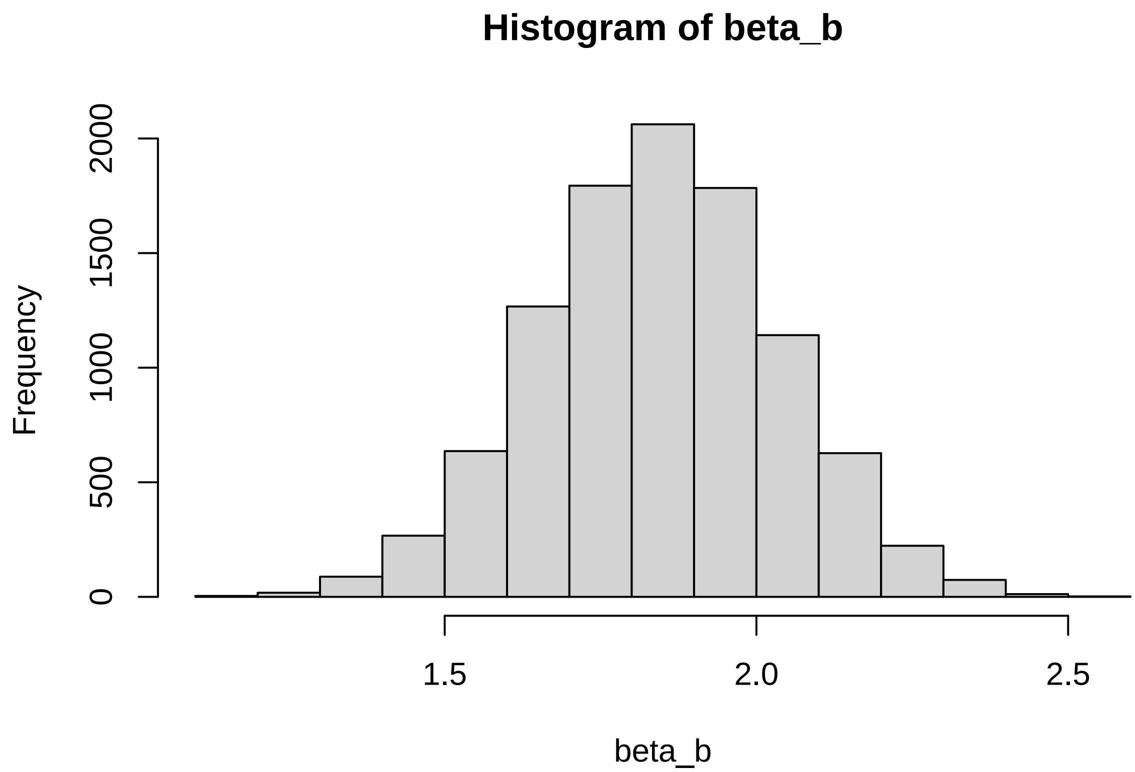
B

```r
sigma2Sample<-BayesLinReg(y, X, mu_0, Omega_0, v_0, sigma2_0, nIter)$sigma2Sample
print(paste0("The posterior median of the standard deviation = ", round(median(sigma2Sample),3)))
```

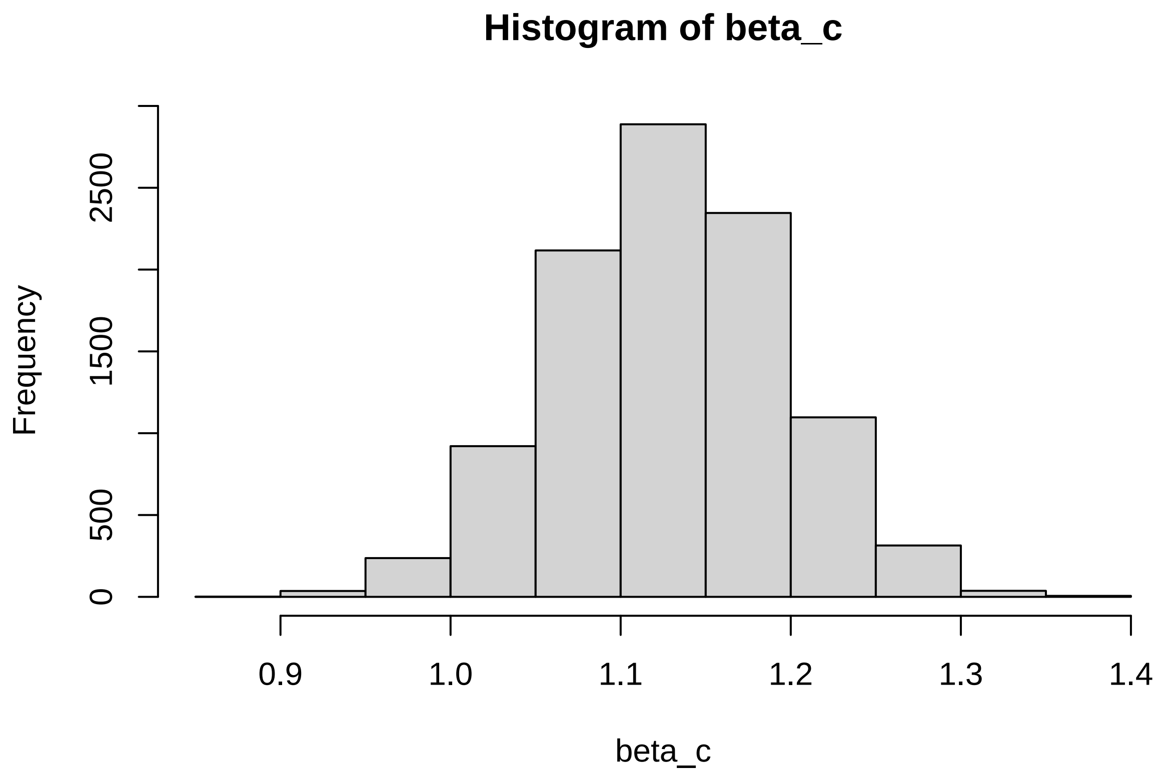
```
## [1] "The posterior median of the standard deviation = 0.408"
```

C

```r
# Solution 1

# We first subset our dataset into 2 , based on the schoole B and C, then we find the beta values by ap
# function in both datasets
hs_B<- subset(X,X[,4]==1)
beta_b<-betaSample%*%t(hs_B)[,2]
# Plot the beta values
hist(beta_b)
```



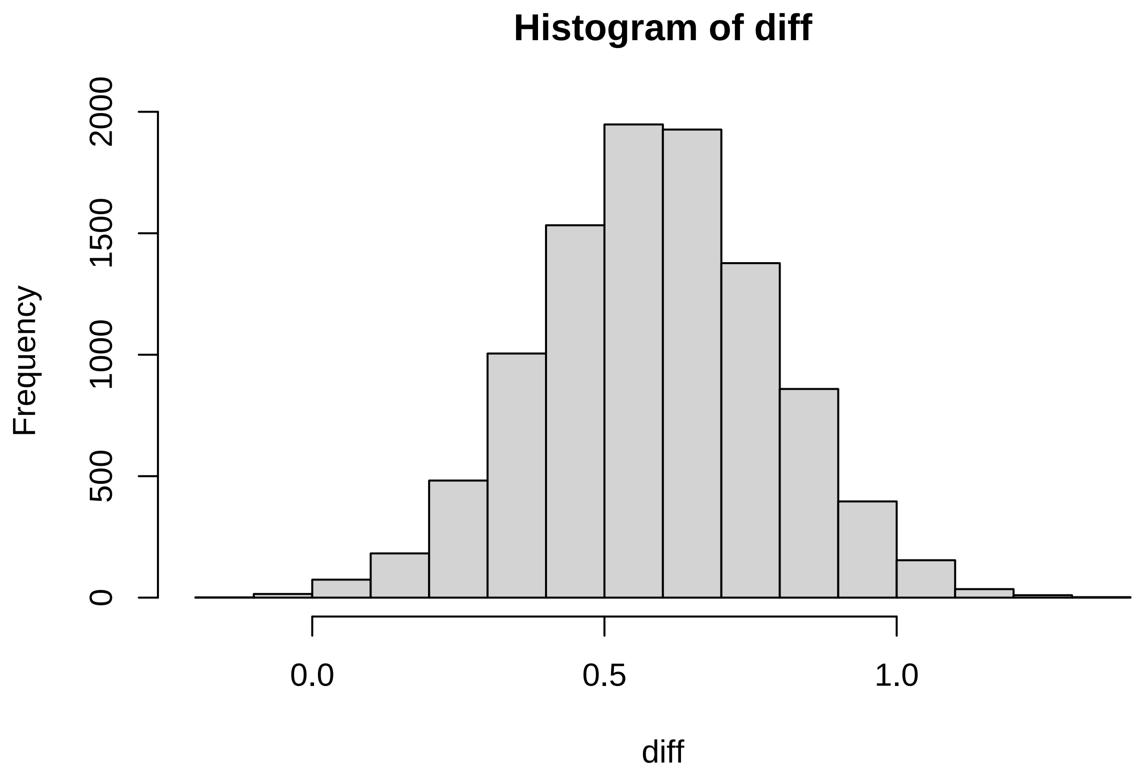

```
hs_C<- subset(X,X[,5]==1)
beta_c<-betaSample%*%t(hs_C)[,2]
# Plot the beta values
hist(beta_c)
```




```
# then we compare the mean value of the of beta_1 for school B vs C
mean(beta_b)/mean(beta_c)-1
```

```
## [1] 0.6334569
```

```
# the results shows that the meain diffrence is .66 whic tells a noticable diff between the
# two schools, however we can back our belive by running one of the indepndent tests to see if this dif

# Solution 2.
# is to sum the effect of beta_1 and the effect if the school for B or C using the betas from the origi
E_B<- betaSample[,2]+betaSample[,4]
E_C<-betaSample[,2]+betaSample[,5]
diff<-E_B-E_C
hist(diff)
```




```
mean(diff)
```

```
## [1] 0.5873029
```

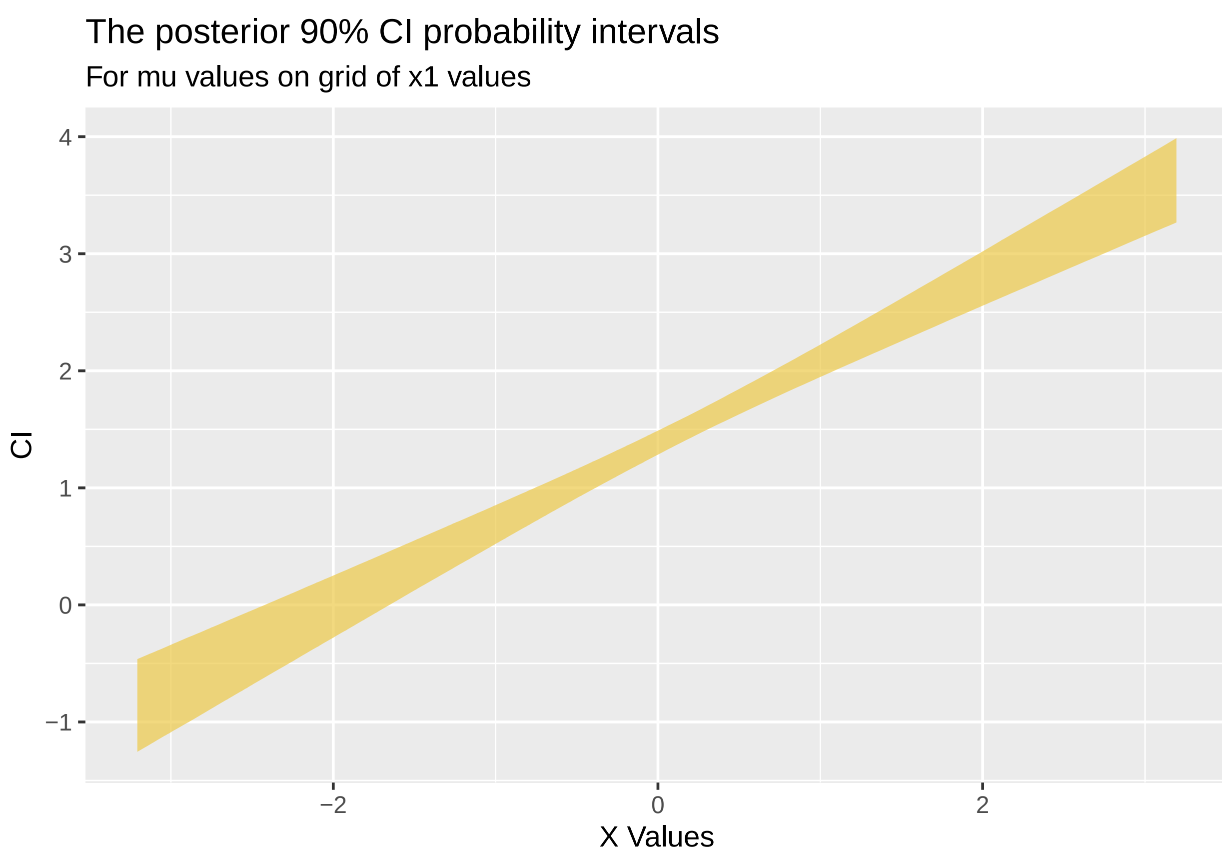
```
quantile(diff,c(.05,.95))
```

```
##        5%       95%
## 0.2597830 0.9168921
```

D

```
const<-1
x1<-seq(min(X[,2]),max(X[,2]),0.01)
x2<-.5
x3<-0
x4<-0
x13<-0
x14<-0

X_new<- as.matrix(data.frame(const=const,x1=x1,x2=x2,x3=x3,x4=x4,x13=x13,x14=x14))

lower<-c()
```

```r
upper<-c()
for (i in 1:nrow(X_new)) {
  mu_val<- betaSample%*%X_new[i,]
  lower[i]<-quantile(mu_val,c(.1))
  upper[i]<-quantile(mu_val,c(.90))
}

plot_df<-data.frame(x=X_new[,2],lower=lower,upper=upper)

ggplot(plot_df, aes(x = x)) +
  geom_ribbon(aes(ymin = lower, ymax = upper)
              , alpha = 0.7,fill = "#EDC948")+
  labs(x = 'X Values', y = 'CI'
       ,title ="The posterior 90% CI probability intervals"
       , subtitle = "For mu values on grid of x1 values"
       ,color = "Line Legend") +
  scale_color_manual(values = c("#14213D","#59A14F","#F28E2B","#EDC948")
                     , labels = c("1","2","3","4"))+
  theme(legend.position="bottom")
```



E

```r
x_new_st<- as.vector(c(1,.4,1,1,0,.4,0))
mu_st<- betaSample %*% x_new_st
e<-c()
for (i in 1:length(sigma2Sample)){
e[i]<-rnorm(1,0,sigma2Sample[1])
}

plt_df2<- data.frame(mu=mu_st,e=e,y=mu_st+e)

ggplot(plt_df2,aes(x = y)) +geom_histogram(aes(y=..density..),
                                             linetype=1, fill='#14213D')+
  geom_density(alpha=.2,color="#FCA311",size=1,fill="#FCA311")+
  labs(x='the posterior predictive distribution of y',y=' ',)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```