Charles Hwang
Dr. Matthews
STAT 488-001
16 June 2022

1 1A) $p(\theta|y) = \text{Beta}(y+1, n-y+1) = \text{Beta}(23+1, 59-23+1) = \boxed{\text{Beta}(24,37)}$

1B) $E(\theta) = \frac{y+1}{n+2} = \frac{23+1}{59+2} = \boxed{\frac{24}{61}}$

1F) After some thought and experimentation, I propose a $\text{Beta}(\frac{23}{59} \cdot 162 + 1, \frac{59-23}{59} \cdot 162 + 1) \approx \boxed{\text{Beta}(64,100)}$ as an informative prior. This is a 23-36 record extrapolated to a full 162-game season (63-99). Using the provided information, $\int_0^{0.25} \text{Beta}(64,100)\, d\theta$ is only ≈ 0.000042 $(\int_{0.75}^{1} \text{Beta}(64,100)\, d\theta = 0)$.

1G) $\text{Beta}(24,37) \cdot \text{Beta}(64,100) =$
$\text{Beta}(24+64-1, 37+100-1) = \boxed{\text{Beta}(87,136)}$

1H) We can see the 95 percent credible interval with the new prior is more narrow, as expected.

2 2A) $p(\mu, \sigma^{-2}|y) = \frac{1}{\sigma^{-n+2}} \frac{1}{e^{2\sigma^{-2}}}((n-1)s^2 + n(\bar{y}-\mu)^2)$

Midterm

Charles Hwang

6/16/2022

Charles Hwang

Dr. Matthews

STAT 488-001

16 June 2022

Problem 1

Problem 1(c)

```
qbeta(0.05/2,23+1,59-23+1)
```

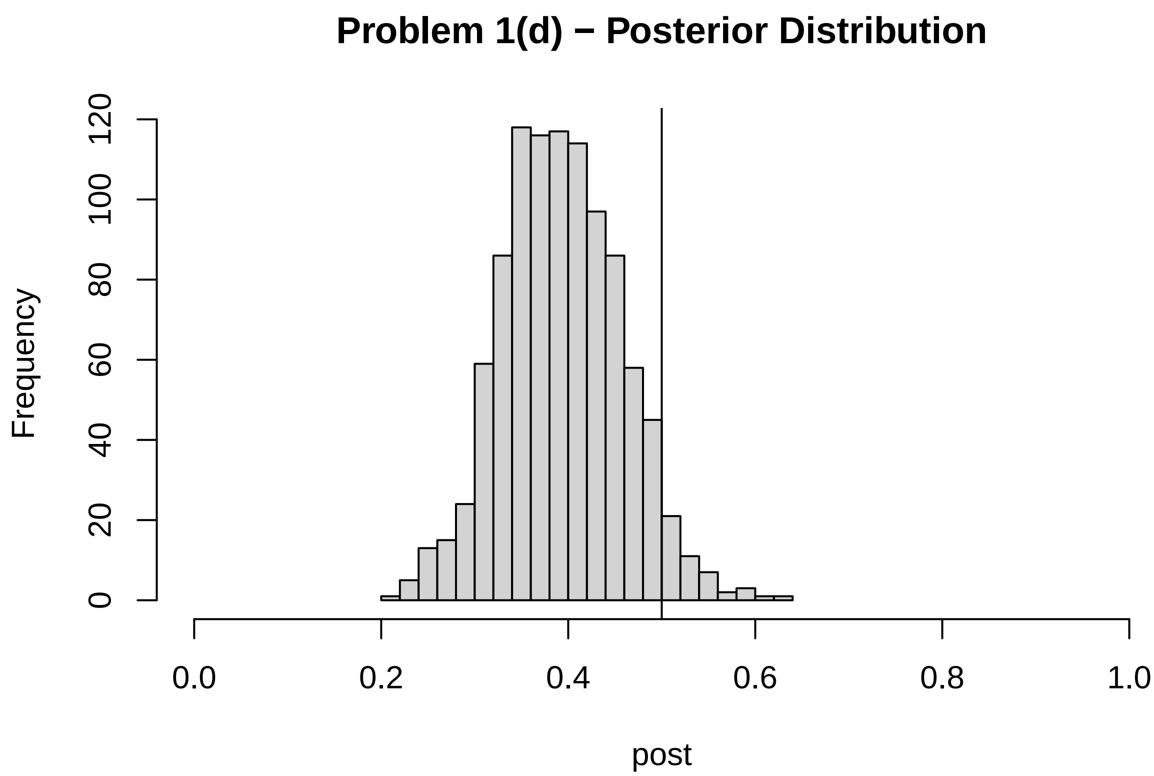
```
## [1] 0.2756216
```

```
qbeta(1-0.05/2,23+1,59-23+1)
```

```
## [1] 0.517885
```

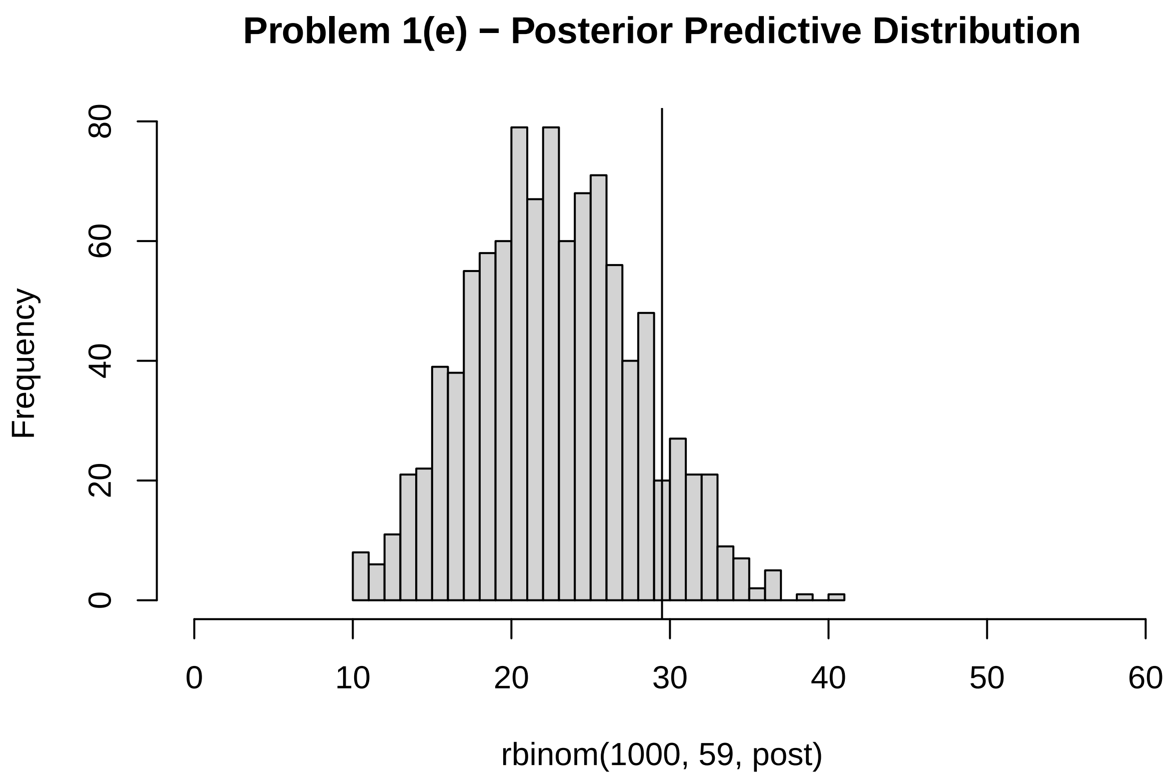
Problem 1(d)

```
set.seed(616)
post<-rbeta(1000,23+1,59-23+1)
hist(post,breaks=20,xlim=c(0,1),main="Problem 1(d) - Posterior Distribution")
abline(v=0.5)
```



Problem 1(e)

```
set.seed(616)
hist(rbinom(1000,59,post),breaks=30,xlim=c(0,59),main="Problem 1(e) - Posterior Predictive Distribution
abline(v=59/2)
```



Problem 1(h)

```
qbeta(0.05/2,23+63+1,59-23+99+1)
```

```
## [1] 0.3272563
```

```
qbeta(1-0.05/2,23+63+1,59-23+99+1)
```

```
## [1] 0.4548796
```

Problem 2

```
p1<-c(69.8,106.9,127.1,41.4,110.7)
p2<-c(162.7,135.6,136.3,135.9,127.7)
```

Problem 2(b)

```
library(extraDistr)
qinvchisq(0.05/2,length(p1)-1,var(p1))
```

```
## [1] 435.558
```

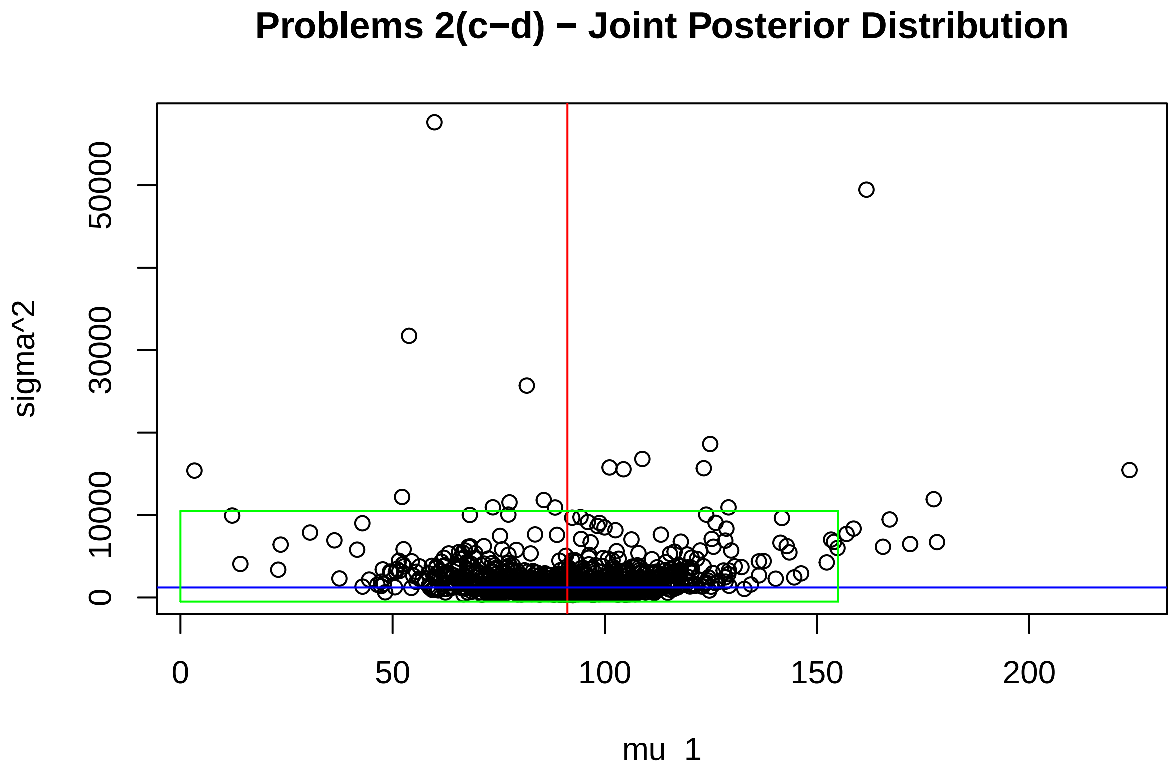
```
qinvchisq(1-0.05/2,length(p1)-1,var(p1))
```

```
## [1] 10019.33
```

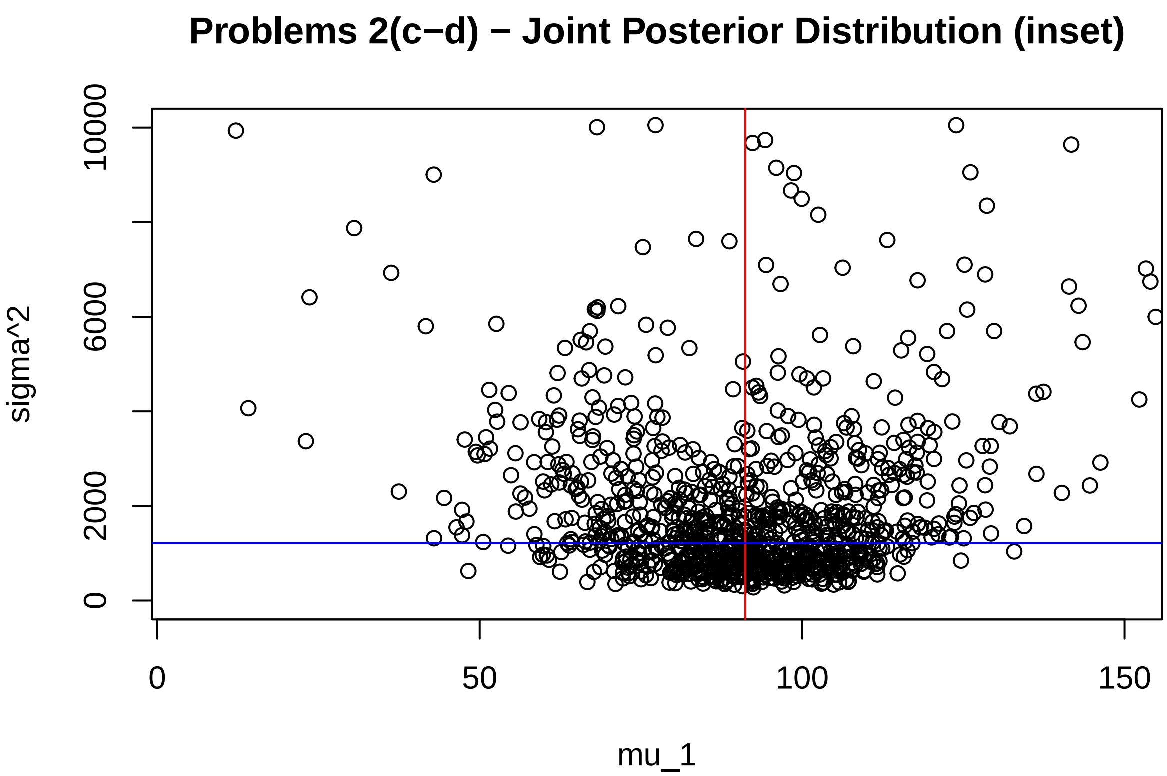
Problems 2(c-d)

```
set.seed(616)
sigma<-rinvchisq(1000,length(p1)-1,var(p1))
```

```r
mu_1<-rnorm(1000,mean(p1),sqrt(sigma/length(p1)))
plot(mu_1,sigma,ylab="sigma^2",main="Problems 2(c-d) - Joint Posterior Distribution") # Problem 2(c)
abline(v=mean(p1),col="red") # Problem 2(d)
abline(h=var(p1),col="blue")
segments(0,10500,0,-500,col="green")
segments(0,10500,155,10500,col="green")
segments(155,10500,155,-500,col="green")
segments(0,-500,155,-500,col="green")
```

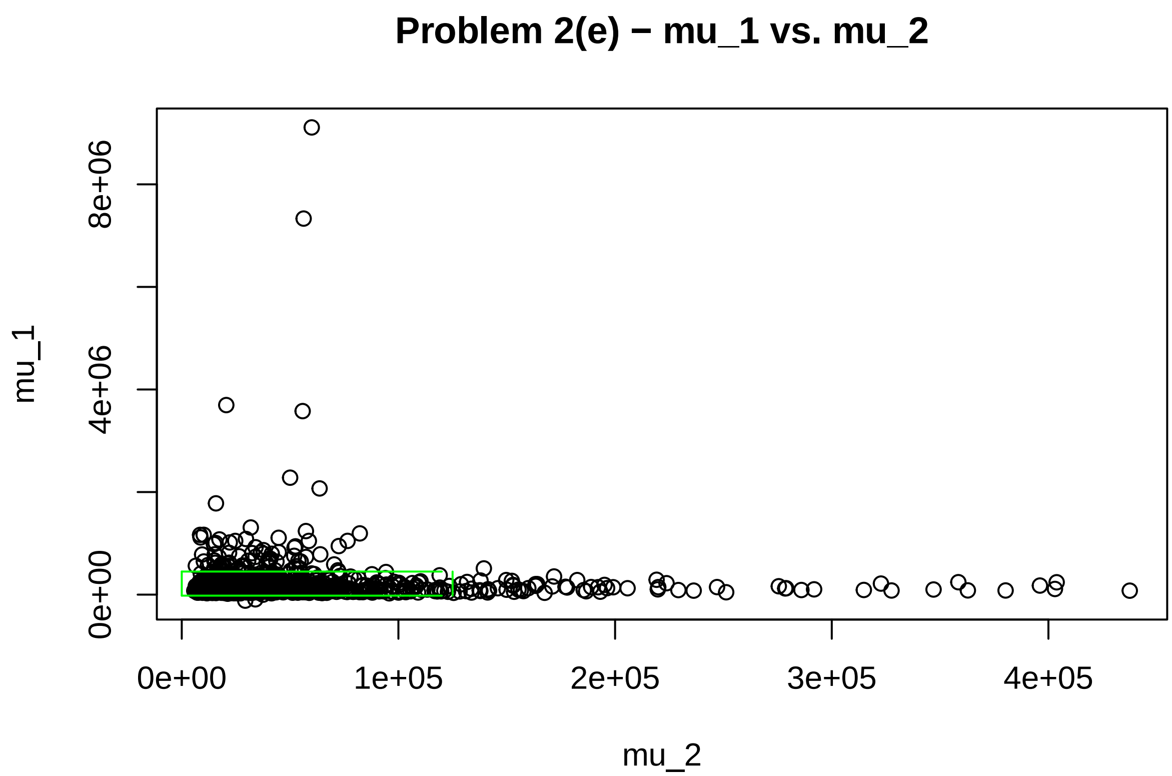



```r
plot(mu_1,sigma,ylab="sigma^2",xlim=c(5,150),ylim=c(0,10000),main="Problems 2(c-d) - Joint Posterior Di
abline(v=mean(p1),col="red") # Problem 2(d)
abline(h=var(p1),col="blue")
```

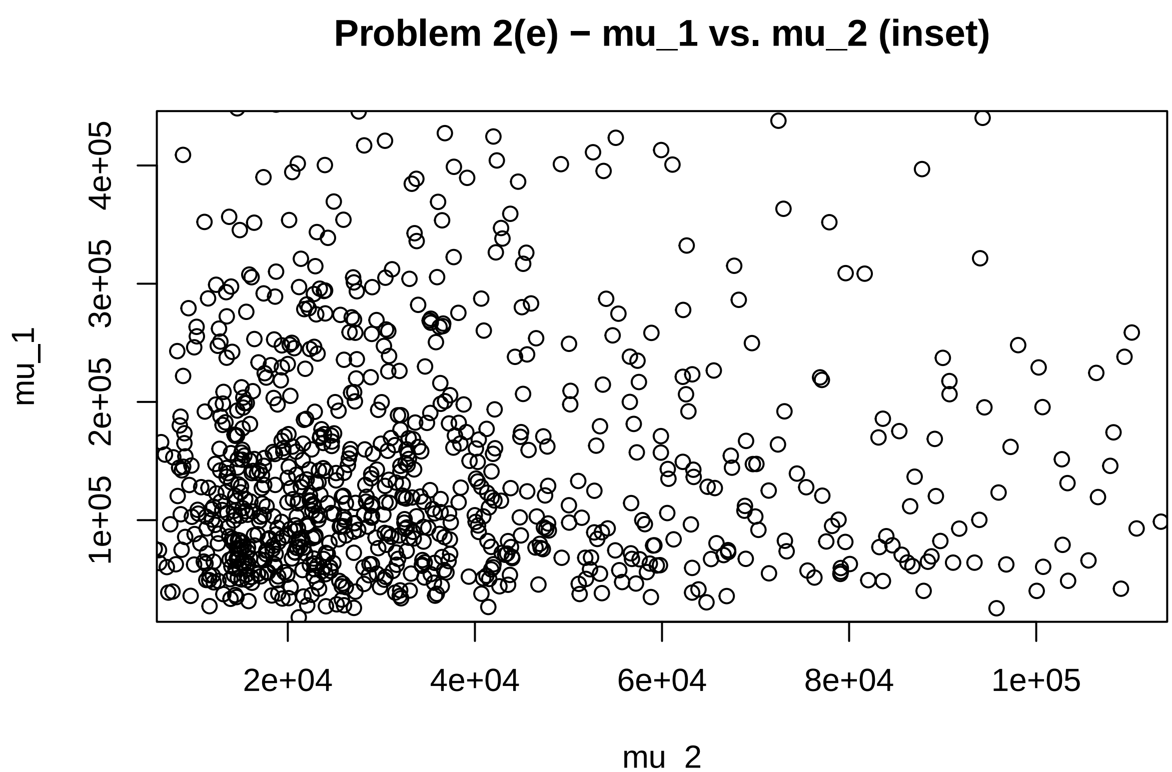


Problem 2(e)

```
set.seed(616)
sig1<-rinvchisq(1000,length(p1)-1,var(p1)) # p(mu,sigma^2|y)=p(mu|sigma^2,y)*p(sigma^2|y)
sig2<-rinvchisq(1000,length(p2)-1,var(p2))
mu_1<-rnorm(1000,mean(p1),sqrt(sig1/length(p1)))
mu_2<-rnorm(1000,mean(p2),sqrt(sig2/length(p2)))
plot(mu_2*sig2,mu_1*sig1,xlab="mu_2",ylab="mu_1",main="Problem 2(e) - mu_1 vs. mu_2")
segments(0,450000,0,-20000,col="green")
segments(0,450000,120000,450000,col="green")
segments(125000,450000,125000,-20000,col="green")
segments(0,-20000,120000,-20000,col="green")
```




```
plot(mu_2*sig2,mu_1*sig1,xlim=c(10000,110000),ylim=c(30000,430000),xlab="mu_2",ylab="mu_1",main="Problem
```



Problem 2(f)

```
mean(p1)-mean(p2)
```

```
## [1] -48.46
```

```
mean(p2)/mean(p1)
```

```
## [1] 1.531476
```

Problem 2(g)

```
sort(mu_1-mu_2)[c(0.05/2,1-0.05/2+1/length(mu_1-mu_2))*length(mu_1-mu_2)]
```

```
## [1] -96.97958  -1.26857
```

Problem 2(h)

```
qinvchisq(0.05/2,length(p1)+length(p2)-1,var(mu_1-mu_2))
```

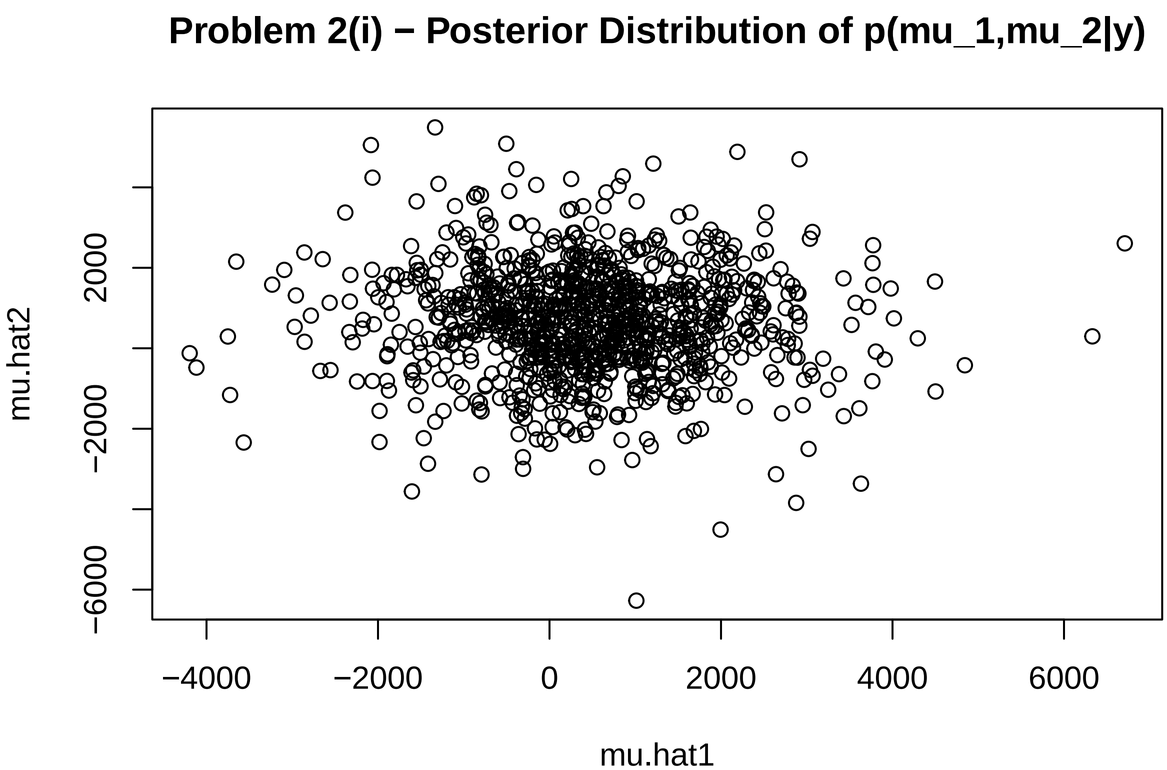
```
## [1] 249.7543
```

```
qinvchisq(1-0.05/2,length(p1)+length(p2)-1,var(mu_1-mu_2))
```

```
## [1] 1759.383
```

Problem 2(i)

```
mu1<-Vectorize(function(t){sum(1/(525+t^2)*p1)/sum(1/(525+t^2))})
pt1<-Vectorize(function(t){sqrt(V(t))*prod((525+t^2)^-0.5*exp(-(p1-mu(t))^2/(2*(525+t^2))))})
mu2<-Vectorize(function(t){sum(1/(525+t^2)*p2)/sum(1/(525+t^2))})
pt2<-Vectorize(function(t){sqrt(V(t))*prod((525+t^2)^-0.5*exp(-(p2-mu(t))^2/(2*(525+t^2))))})
mu<-Vectorize(function(t){sum(1/(525+t^2)*c(p1,p2))/sum(1/(525+t^2))})
pt<-Vectorize(function(t){sqrt(V(t))*prod((525+t^2)^-0.5*exp(-(c(p1,p2)-mu(t))^2/(2*(525+t^2))))})
V<-Vectorize(function(t){1/(sum(1/(525+t^2)))})
tau<-seq(1,3000,1)
set.seed(616)
V_mu<-sample(V(tau),1000,replace=TRUE,prob=pt(tau))
mu.hat1<-rnorm(1000,mu1(tau),sqrt(V_mu))
mu.hat2<-rnorm(1000,mu2(tau),sqrt(V_mu))
plot(mu.hat1,mu.hat2,main="Problem 2(i) - Posterior Distribution of p(mu_1,mu_2|y)")
```



Problem 2(j)

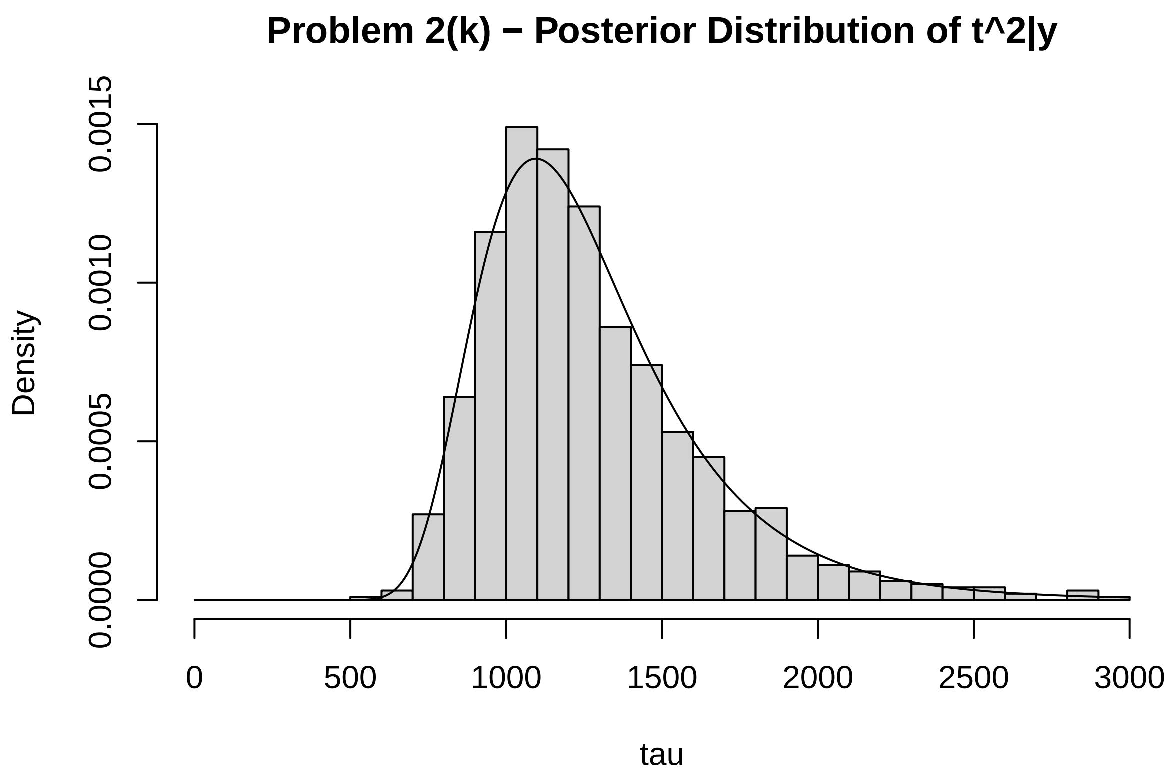
```
sort(mu.hat1-mu.hat2)[c(0.05/2,1-0.05/2+1/length(mu.hat1-mu.hat2))*length(mu.hat1-mu.hat2)]
```

```
## [1] -4073.283  3743.447
```

We can see that this 95 percent credible interval is clearly much wider than the previous one (-96.9795791, -1.2685701) from problem 2(g).

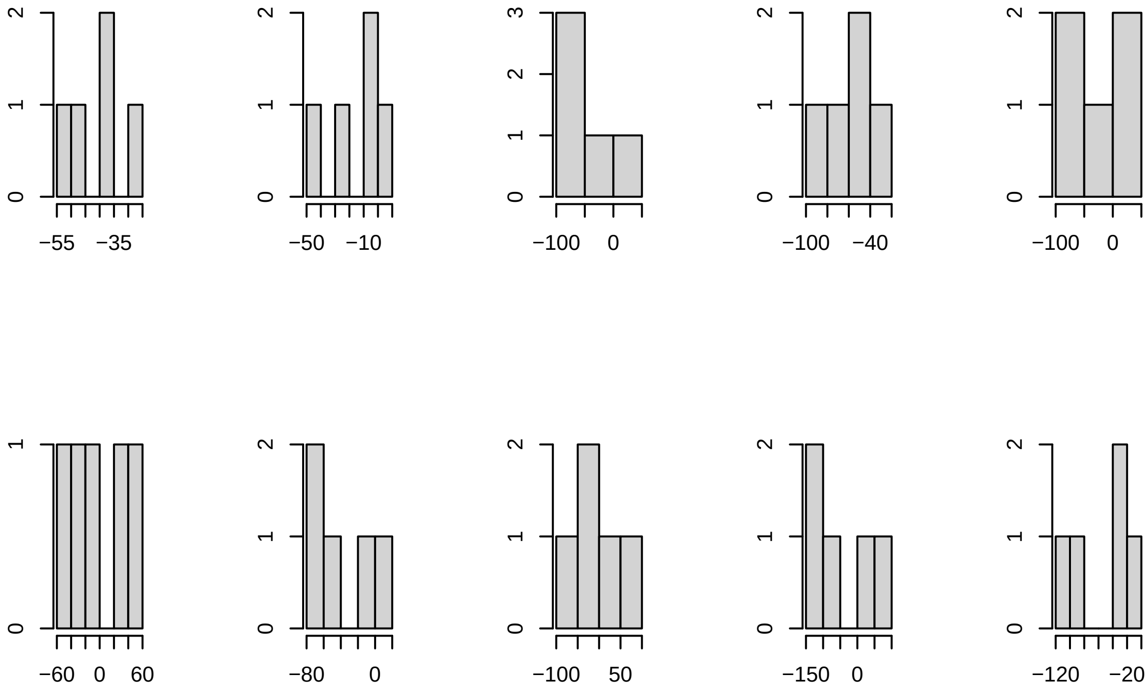
Problem 2(k)

```
set.seed(616)
ty<-sample(tau,1000,replace=TRUE,prob=pt(tau))
hist(ty,breaks=30,freq=FALSE,xlim=c(0,3000),xlab="tau",main="Problem 2(k) - Posterior Distribution of t
points(tau,pt(tau)/sum(pt(tau)),type="l")
```

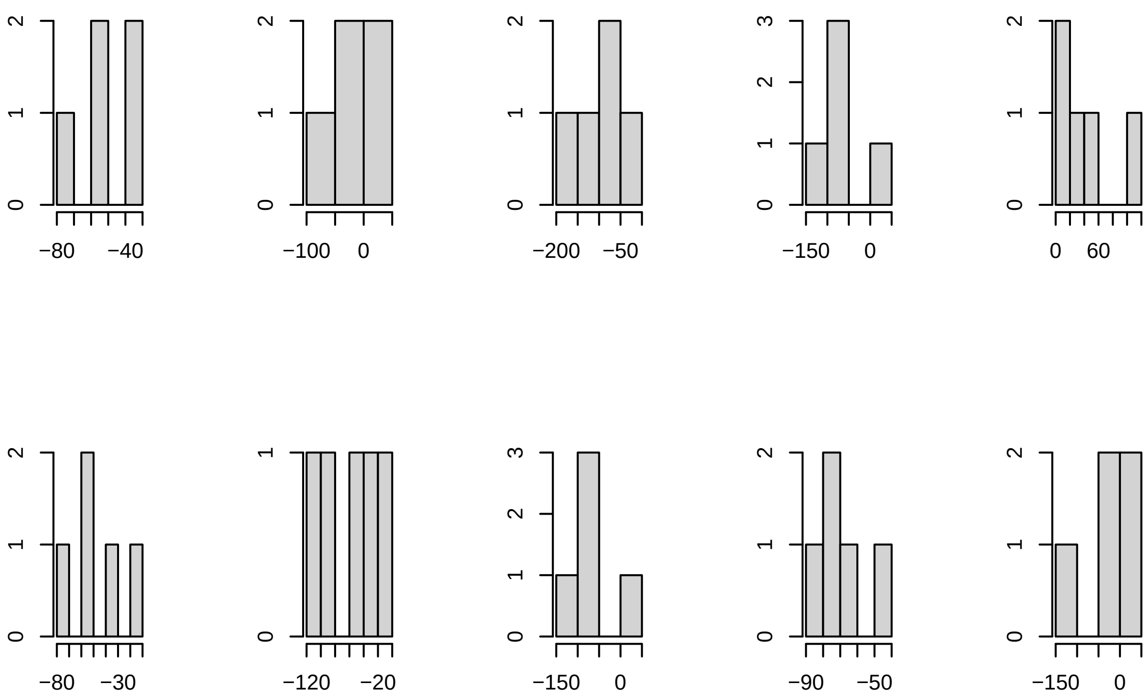


Problem 2(l)

```
set.seed(616)
sig1<-rinvchisq(20,length(p1)-1,var(p1))
sig2<-rinvchisq(20,length(p2)-1,var(p2))
mu_1<-rnorm(20,mean(p1),sqrt(sig1/length(p1)))
mu_2<-rnorm(20,mean(p2),sqrt(sig2/length(p2)))
par(mfrow=c(2,5))
hist(rnorm(length(p1),mu_1[1]-mu_2[1],sqrt(sig1[1])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[2]-mu_2[2],sqrt(sig1[2])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[3]-mu_2[3],sqrt(sig1[3])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[4]-mu_2[4],sqrt(sig1[4])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[5]-mu_2[5],sqrt(sig1[5])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[6]-mu_2[6],sqrt(sig1[6])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[7]-mu_2[7],sqrt(sig1[7])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[8]-mu_2[8],sqrt(sig1[8])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[9]-mu_2[9],sqrt(sig1[9])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[10]-mu_2[10],sqrt(sig1[10])),xlab="",ylab="",main="")
```




```
hist(rnorm(length(p1),mu_1[11]-mu_2[11],sqrt(sig1[11])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[12]-mu_2[12],sqrt(sig1[12])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[13]-mu_2[13],sqrt(sig1[13])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[14]-mu_2[14],sqrt(sig1[14])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[15]-mu_2[15],sqrt(sig1[15])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[16]-mu_2[16],sqrt(sig1[16])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[17]-mu_2[17],sqrt(sig1[17])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[18]-mu_2[18],sqrt(sig1[18])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[19]-mu_2[19],sqrt(sig1[19])),xlab="",ylab="",main="")
hist(rnorm(length(p1),mu_1[20]-mu_2[20],sqrt(sig1[20])),xlab="",ylab="",main="")
```



It is difficult to tell with so many histograms and such a small sample size, but we can see the histograms appear to be roughly normal. All histograms overlap at least partially with the true value of $\mu_1 - \mu_2$ from our data.