Homework 8

Shao-Ting Chiu (UIN:433002162)

11/8/22

Table of contents

Description

- Course: STAT638, 2022 Fall

  Read Chapter 8 in the Hoff book. Then do the following exercises in Hoff: 8.1 and 8.3.

  Please note some typos in 8.1: All θ_i's should be θ_j's.

  For 8.1(c), you may find the law of total (co-)variance useful. In addition, remember that all of these laws also hold for conditional distributions (e.g., when conditioning on additional quantities such as μ and τ^2 in all terms on the left- and right-hand side of the equation).

Computational Enviromnent Setup[1]

Third-party libraries

```julia
using Pkg
Pkg.activate("hw8")
using Distributions
using DataFrames
using Turing
using Plots
using DelimitedFiles
using LinearAlgebra
using Statistics
using Turing
using StatsBase
using StatsPlots
import Random
Random.seed!(2022)
```

```
  Activating project at `~/Documents/GitHub/STAT638_Applied-Bayes-Methods/hw/hw8`
```

```
Random.TaskLocalRNG()
```

Version

```julia
Pkg.status()
VERSION
```

```
Status `~/Documents/GitHub/STAT638_Applied-Bayes-Methods/hw/hw8/Project.toml`
  [a93c6f00] DataFrames v1.4.2
  [31c24e10] Distributions v0.25.76
  [be115224] MCMCDiagnosticTools v0.1.4
 [91a5bcdd] Plots v1.35.5
  [2913bbd2] StatsBase v0.33.21
  [f3b207a7] StatsPlots v0.15.4
  [fce5fe82] Turing v0.21.12
  [8bb1440f] DelimitedFiles
```

[1] I use special character in Julia code. Unfortunately, those are not displayed in PDF version.

```
  [10745b16] Statistics
Info Packages marked with  have new versions available and may be upgradable.

v"1.8.2"
```

Problem 8.1

Components of variance: Consider the hierarchical model where

$$\theta_1, \dots, \theta_m | \mu, \tau^2 \sim i.i.d.\text{normal}(\mu, \tau^2)$$

$$y_{1,j}, \dots, y_{n_j,j} | \theta_j, \sigma^2 \sim i.i.d.\text{normal}(\theta_j, \sigma^2)$$

For this problem, we will eventually compute the following:

- $Var[y_{i,j}|\theta_i, \sigma^2]$, $Var[\bar{y}_{\cdot,j}|\theta_i, \sigma^2]$, $Cov[y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2]$
- $Var[y_{i,j}|\mu, \tau^2]$, $Var[\bar{y}_{\cdot,j}|\mu, \tau^2]$, $Cov[y_{i_1,j}, y_{i_2,j}|\mu, \tau^2]$ First, lets use our intuition to guess at the answers:

(a)

Which do you think is bigger, $Var[y_{i,j}|\theta_i, \sigma^2]$ or $Var[y_{i,j}|\mu, \tau^2]$? To guide your intuition, you can interpret the first as the variability of the Y's when sampling from a fixed group, and the second as the variability in first sampling a group, then sampling a unit from within the group.

- $Var[y_{i,j}|\mu, \tau^2]$ because θ_j is uncertain and the between-group varibability create additional uncertainty.

(b)

Do you think $Cov[y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2]$ is negative, positive, or zero? Answer the same for $Cov[y_{i_1,j}, y_{i_2,j}|\mu, \tau^2]$. You may want to think about what $y_{i_2,j}$ tells you about $y_{i_1,j}$ if θ_j is known, and what it tells you when θ_j is unknown.

$Cov[y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2]$

Because $y_{i_1,j}$ and $y_{i_2,j}$ is i.i.d. sampled, I expect $Cov[y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2]$ to be zero.

$Cov[y_{i_1,j}, y_{i_2,j}|\mu, \tau^2]$

$y_{1,j}$ does tell information about $y_{2,j}$. The covariance $Cov[y_{i_1,j}, y_{i_2,j}|\mu, \tau^2]$ is likely to be positive because values from same θ_j tend to be close together.

(c)

> Now compute each of the six quantities above and compare to your answers in (a) and (b). [2]

$$Var[y_{i,j}|\theta_i, \sigma^2] = \sigma^2 \tag{1}$$

$$\begin{aligned}
Var[\bar{y}_{\cdot,j}|\theta_i, \sigma^2] &= Var[\sum_{i'=1}^{n_j} y_{i',j}/n|\theta_i, \sigma^2] &(2)\\
&= \frac{1}{n^2} Var[\sum_{i'=1}^{n_j} y_{i',j}|\theta_i, \sigma^2] &(3)\\
&= \frac{1}{n^2} \sum_{i'=1}^{n_j} Var[y_{i',j}|\theta_i, \sigma^2] &(4)\\
&= \frac{1}{n} Var[y_{i',j}|\theta_i, \sigma^2] &(5)\\
&= \frac{\sigma^2}{n} &(6)
\end{aligned}$$

$$\begin{aligned}
Cov[y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2] &= E[y_{i_1,j} y_{i_2,j}] - E[y_{i_1,j}]E[y_{i_2,j}] &(7)\\
&= E[y_{i_1,j}]E[y_{i_2,j}] - E[y_{i_1,j}]E[y_{i_2,j}] &(8)\\
&= 0 &(9)
\end{aligned}$$

[2] $Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$

$$
\begin{aligned}
Var[y_{i,j}|\mu,\tau^2] &= E(Var[y_{i,j}|\mu,\tau^2,\theta,\sigma^2]|\mu,\tau^2) + Var(E[y_{i,j}|\mu,\tau^2,\theta,\sigma^2]|\mu,\tau^2) &(10)\\
&= E(\sigma^2|\mu,\tau^2) + Var(\theta|\mu,\tau^2) &(11)\\
&= \sigma^2 + \tau^2 &(12)
\end{aligned}
$$

$$
\begin{aligned}
Var[\bar{y}_{\cdot,j}|\mu,\tau^2] &= E(Var[\bar{y}_{\cdot,j}|\mu,\tau^2,\theta,\sigma^2]|\mu,\tau^2) + Var(E[\bar{y}_{\cdot,j}|\mu,\tau^2,\theta,\sigma^2]|\mu,\tau^2) &(13)\\
&= E(\frac{\sigma^2}{n}|\mu,\tau^2) + Var(\theta|\mu,\tau^2) &(14)\\
&= \frac{\sigma^2}{n} + \tau^2 &(15)
\end{aligned}
$$

$$
\begin{aligned}
Cov[y_{i_1,j}, y_{i_2,j}|\mu,\tau^2] &= E(Cov[y_{i_1,j}, y_{i_2,j}|\theta,\sigma^2,\mu,\tau^2]|\mu,\tau^2) &(16)\\
&\quad + Cov(E[y_{i_1,j}|\theta,\sigma^2,\mu,\tau^2], E[y_{i_2,j}|\theta,\sigma^2,\mu,\tau^2]|\mu,\tau^2) &(17)\\
&= 0 + Cov(\theta,\theta|\mu,\tau^2) &(18)\\
&= E[\theta^2|\mu,\tau^2] - E[\theta|\mu,\tau^2]^2 &(19)\\
&= Var(\theta|\mu,\tau^2) &(20)\\
&= \tau^2 &(21)
\end{aligned}
$$

(d)

Now assume we have a prior $p(\mu)$ for μ. Using Bayes' rule, show that

$$p(\mu|\theta_1,\dots,\theta_m,\sigma^2,\tau^2,y_1,\dots,y_m) = p(\mu|\theta_1,\dots,\theta_m,\tau^2)$$

Interpret in words what this means.

$$
\begin{aligned}
p(\mu|\theta_1,\dots,\theta_m,\sigma^2,\tau^2,y_1,\dots,y_m) &= \frac{p(\sigma^2,y_1,\dots,y_m|\mu,\theta_1,\dots,\theta_m,\tau^2)p(\mu|\theta_1,\dots,\theta_m,\tau^2)}{p(\sigma^2,y_1,\dots,y_m|\theta_1,\dots,\theta_m,\tau^2)} &(22)\\
&= p(\mu|\theta_1,\dots,\theta_m,\tau^2) &(23)
\end{aligned}
$$

where $p(\sigma^2,y_1,\dots,y_m|\mu,\theta_1,\dots,\theta_m,\tau^2) = p(\sigma^2,y_1,\dots,y_m|\theta_1,\dots,\theta_m,\tau^2)$ because knowing μ doesn't provide more information when $\theta_1,\dots,\theta_m$ are known.

Problem 8.3

Herarchical modeling: The files `school1.dat` through `school8.dat` give weekly hours spent on homework for students sampled from eight different schools. Obtain posterior distributions for the true means for the eight different schools using a herarchical normal model with the following prior parameters:

$$\mu_0 = 7, \gamma_0^2 = 5, \tau_0^2 = 10, \eta_0 = 2, \sigma_0^2 = 15, \nu_0 = 2$$

```
dsch = Dict()
nsch = 8
for i in 1:nsch
   dsch[i] = readdlm("data/school$i.dat")
end
```

(a)

Run a Gibbs sampling algorithm to approximate the posterior distribution of $\{\theta, \sigma^2, \mu, \tau^2\}$. Assess the convergence of the Markov chain, and find the effective sample size for $\{\sigma^2, \mu, \tau^2\}$. Run the chain long enough so that the effective sample sizes are all above 1000.

```
# Prior
 0 = 7.
 0² = 5.
 0² = 10.
 0 = 2.
 0² = 15.
 0 = 2.

# Data
ns = [ length(dsch[i]) for i in 1:nsch]
n = sum(ns)
m = length(dsch)
ȳs = [mean(dsch[i]) for i in 1:nsch]
s²s = [ (ns[i] - 1)^-1 * sum( (dsch[i] .- ȳs[i]).^2) for i in 1:nsch]
# posterior

function  ²_pos(m,  0, v,  ,  0²)
    ths² = sum([ ( -  )^2 for   in  v])
      = (m +  0)* 0.5
```

```
      = (ths² +  0* 0²)/2
    return InverseGamma( ,  )
end

function  ²_pos(n,  0,  0², ns, s²s, ȳs,  s)
      = (n+ 0)/2
      =( sum((ns .- 1) .* s²s .+ ns .* (ȳs .-  s).^2) +  0* 0²)/2
    return InverseGamma( ,  )
end

function  _pos(m,  ²,  s,  0²,  0)
     m² = (m/ ² + 1/ 0²)^-1
    ¯ = mean( s)

    a =  m²*(m*¯/ ² +  0/ ²)
    return Normal(a,  m²)
end

function  _pos( ², ȳ, n,  ²,  )
    ~² = (n/ ² + 1/ ²)^-1
    a = ~²*(n*ȳ/ ² +  / ²)
    return Normal(a, ~²)
end

"""
Effective Sample Size
"""
function ess(v)
    n = length(v)
    c = sum(autocov(v, collect(1:n-1)))
    return n/(1+2*c)
end

# Sampling
smp = 4000
 ²s = zeros(smp)
 ²s = zeros(smp)
 s = zeros(smp)
 s = zeros(smp, m)
```

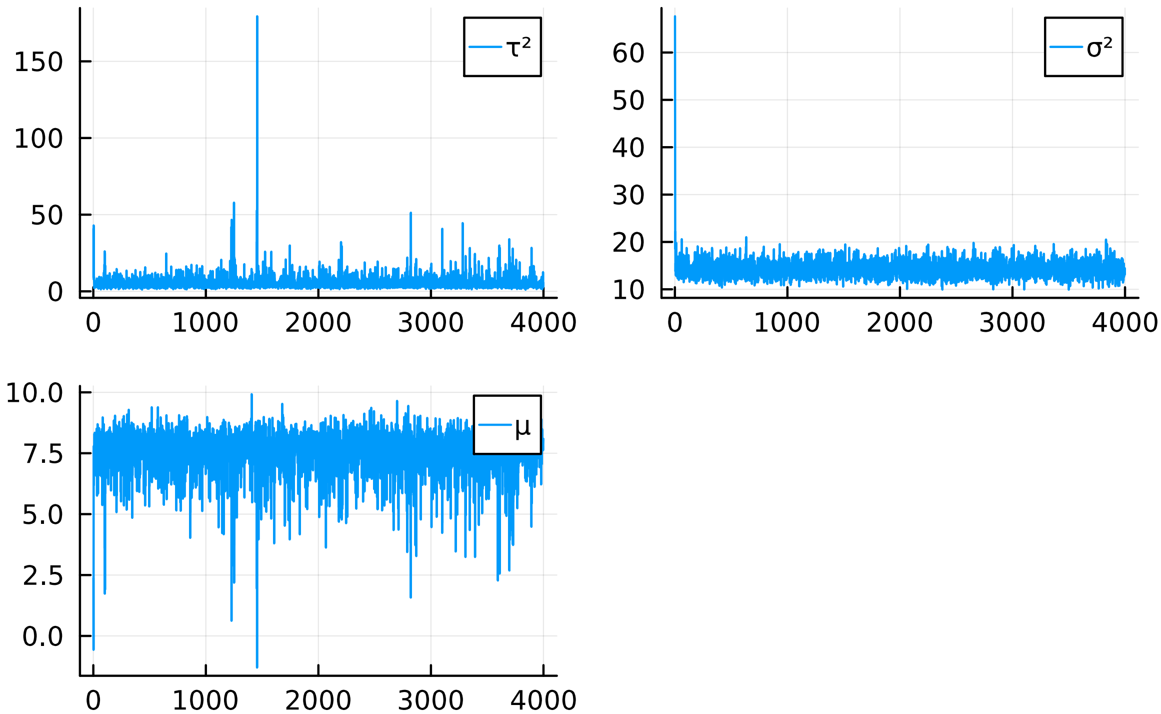
```
 ²s[1] = rand(InverseGamma( 0/2,  0∗ 0²/2))
 ²s[1] = rand(InverseGamma( 0 /2,  0 ∗ 0²/2))
 s[1] = rand(Normal( 0,  0²))
# s[1,:] = [rand( _pos( ²s[1], ȳs[i], ns[i],  ²s[1],  s[1])) for i in 1:m]
 s[1,:] = rand(Normal( s[1],  ²s[1]), m)

for s in 2:smp
     ²s[s] = rand( ²_pos(n,  0,  0², ns, s²s, ȳs,  s[s-1,:]))
     ²s[s] = rand( ²_pos(m,  0,  s[s-1,:],  s[s-1],  0²))
     s[s,:] = [rand( _pos( ²s[s-1], ȳs[i], ns[i],  ²s[s-1],  s[s-1])) for i in 1:m]
     s[s] = rand( _pos(m,  ²s[s-1],  s[s-1,:],  0²,  0))
end
```

```
for i in [ ²s,  ²s,  s,  s]
    plot(i)
end

p1 = plot( ²s[2:end], label=" ²")
p2 = plot( ²s[2:end], label=" ²")
p3 = plot( s[2:end], label=" ")

plot(p1, p2, p3)
```



- Effetive Sample Size
- τ^2

```
ess( ²s)
```

```
1523.525
```

- σ^2

```
ess( ²s)
```

```
1234.234
```

- μ

```
ess( s)
```

```
1045.242
```

(b)

> Compute posterior means and 95% confidence regions for $\{\sigma^2, \mu, \tau^2\}$. Also, compare the posterior densities to the prior densities, and discuss what was learned from the data.

- σ^2

```
quantile( ²s, [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 11.492557006363636
 14.15210768607951
 17.713837318481733
```

- μ

```
quantile( s, [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 5.3131625605052655
 7.688237183428129
 8.736460757511468
```

- τ

```
quantile( ²s, [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
  1.8676028002352072
  4.466482143729323
 15.698240842165873
```

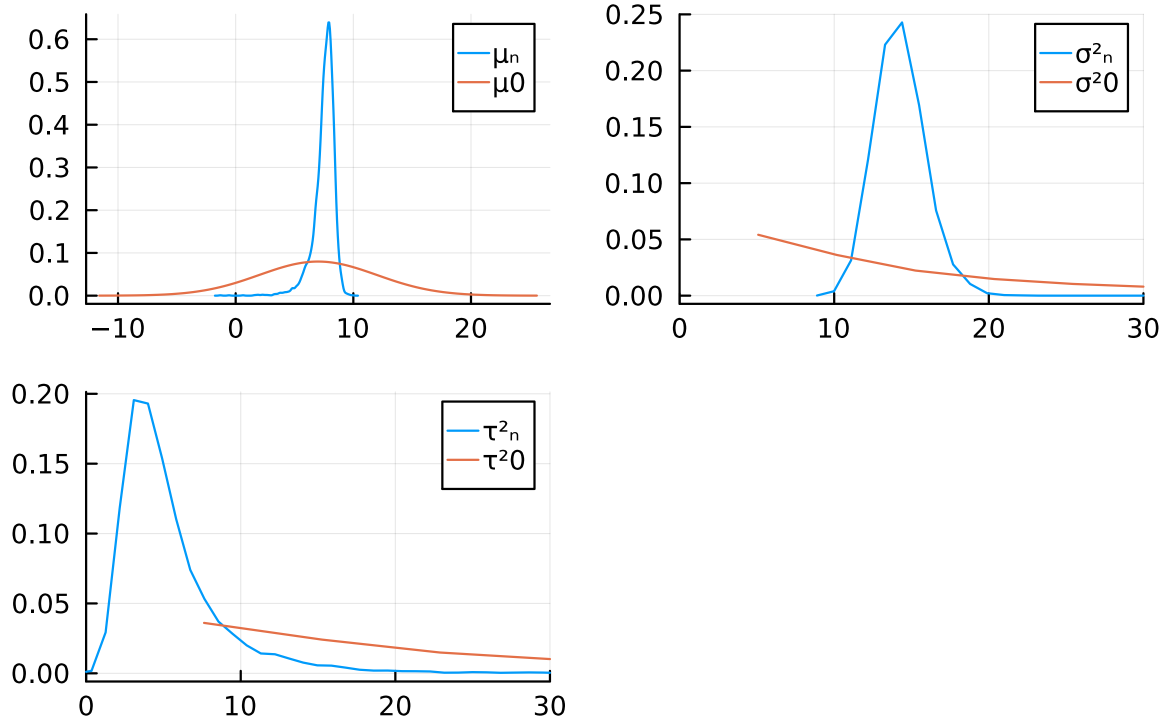
```
pu = density( s, label=" ")
pt = density( ²s,label=" ² ")
ps = density( ²s, label=" ² ")

plot!(pu, Normal( 0 ,  0²), label= " 0")
plot!(pt, InverseGamma( 0 /2,  0 * 0²/2), label= " ²0")
```

```
plot!(ps, InverseGamma( 0/2,  0* 0²/2), label= " ²0")

xlims!(pt, 0,30)
xlims!(ps, 0,30)

plot(pu, pt, ps)
```



Estimations of μ and τ are similar in prior and posterior. However, σ^2 is different.

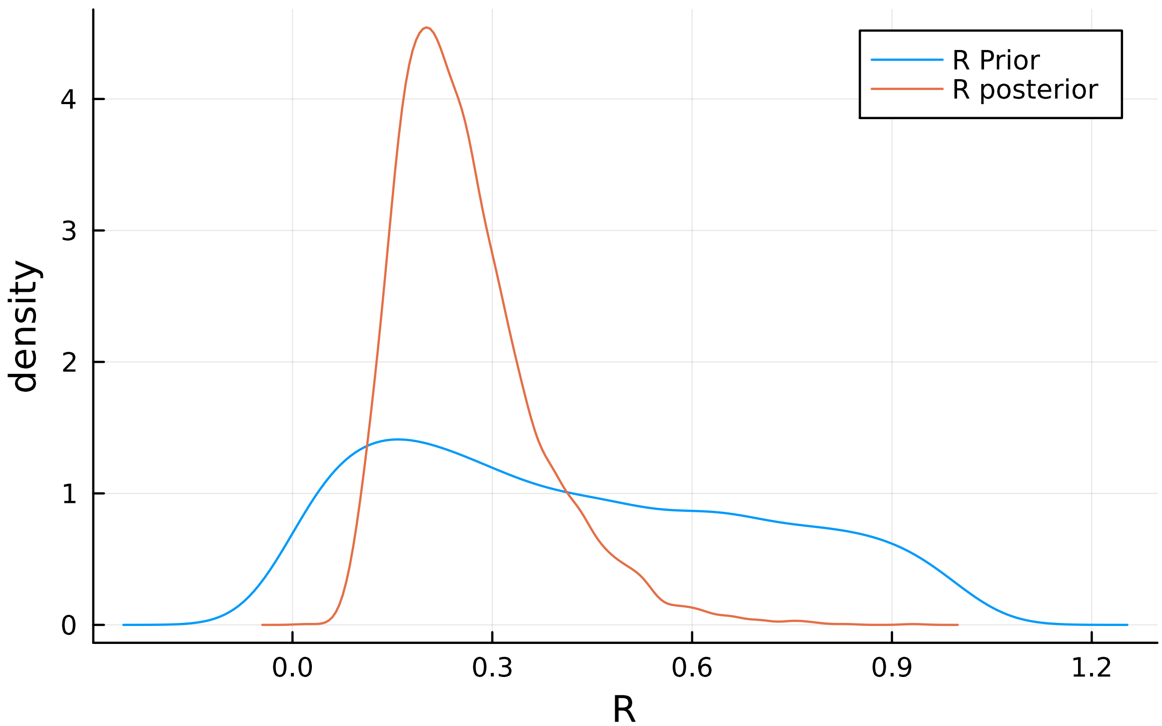
(c)

Plot the posterior density of $R = \frac{\tau^2}{\sigma^2+\tau^2}$ and compare it to a plot of the prior density of R. Describe the evidence for between-school variation.

```
²_prs = rand(InverseGamma( 0/2,  0* 0²/2), 1000)
²_prs = rand(InverseGamma( 0 /2,  0 * 0²/2), 1000)
```

```
R_prs =  ²_prs ./ ( ²_prs .+  ²_prs)
R_pos =  ²s ./ ( ²s .+  ²s)

pr = density(R_prs, label="R Prior", xlabel="R", ylabel="density")
density!(pr, R_pos, label="R posterior")
```



R represents the quantity of vairance in between-group. The prior is not certain about the specific quantity, but after applying posterior inference. The posterior probability of R is peaked and more cetain about the value is around 0.3.

(d)

> Obtain the posterior probability that θ_7 is smaller than θ_6, as well as the posterior probability that θ_7 is smaller than of all the θ's.

- p(θ_7 is smaller than θ_6)

```
mean( s[:,7] .<  s[:,6])
```

```
0.53075
```

- p(θ_7 is smaller than of all the θ's)

```
res = zeros(size( s)[1])
for i in 1 : size( s)[1]
    if argmin( s[i,:]) == 7
        res[i] = 1
    end
end
mean(res)
```

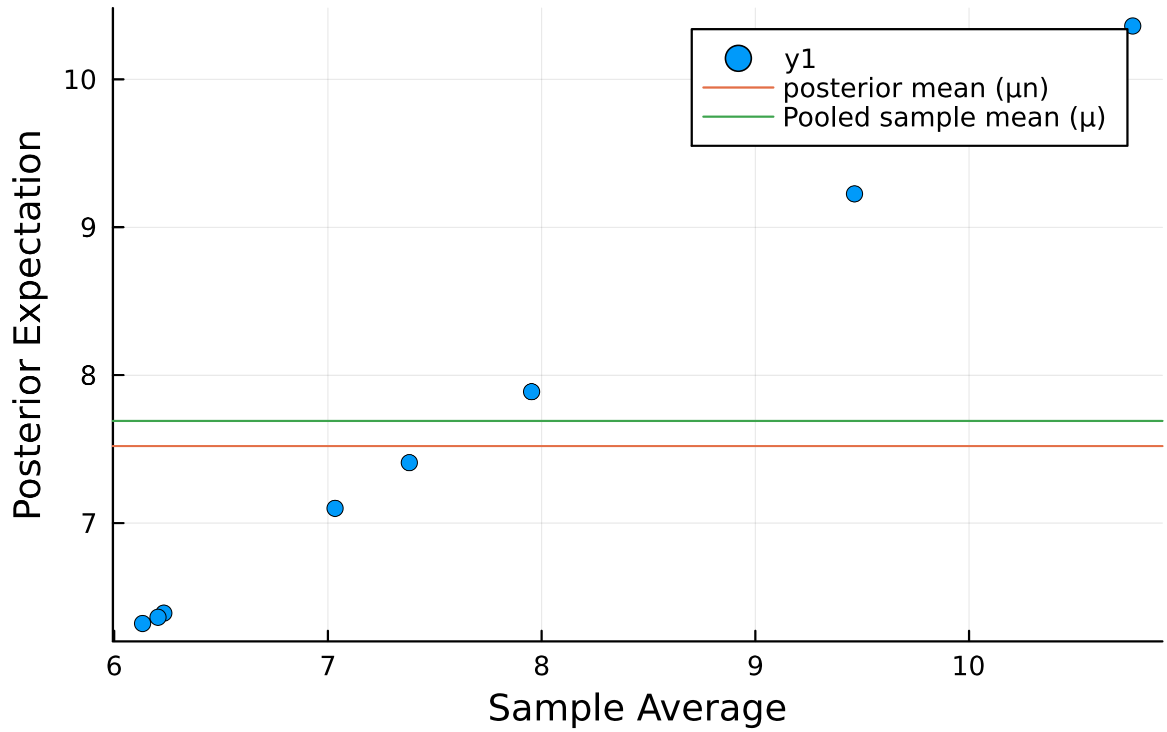
```
0.339
```

(e)

> Plot the sample averages $\bar{y}_1, \ldots, \bar{y}_8$ against the posterior expectations of $\theta_1, \ldots, \theta_8$, and describe the relationship. Also compute the sample mean of all observations and compare it to the posterior mean of μ.

```
psmp = scatter(ȳs, mean( s, dims = 1)[1,:], xlabel="Sample Average", ylabel= "Posterior Exp

hline!(psmp, [mean( s)], label="posterior mean ( n)")
hline!(psmp, [sum(ȳs .* ns)/n], label="Pooled sample mean ( )" )
```


y1
posterior mean (μn)
Pooled sample mean (μ)
Posterior Expectation
Sample Average
10
9
8
7
6
7
8
9
10