ROS_CHP9

Michael Rowlands

2024-04-13

Question 7

Repeat analysis of section 9.5 on the earnings data - that is, perform a mock sensitivity analysis on the choice of prior on the inference of the realationship between height and earnings.

We will compare the inferences of the relationship between centered height (height - sample_mean(height) in inches) and earnings (dollars/year) coming from a linear model of:

$$earnings_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta(\text{height}_i - \bar{\text{height}})$$

with the choice of priors being imporper as in:

$$\alpha \sim \text{Unif}(-\infty, \infty)$$

$$\beta \sim \text{Unif}(-\infty, \infty)$$

Weak default priors as in:

$$\alpha \sim \text{Normal}(\bar{\text{earnings}}, 2.5 * \text{sd}(\text{earnings}))$$
$$\beta \sim \text{Normal}\left(0, 2.5 * \frac{\text{sd}(\text{earnings})}{\text{sd}(\text{height})}\right)$$

Weakly Informative priors as in:

$$\alpha \sim \text{Normal}(20000, 2500)$$
$$\beta \sim \text{Normal}\,(0, 1000)$$

We will also consider the results of being too restrictive on the prior for the β coefficent as in:

$$\alpha \sim \text{Normal}(20000, 2500)$$
$$\beta \sim \text{Normal}\,(0, 250)$$

```
#Q7

data(earnings)

#Center height for easier intercept prior
earnings$c_height <- earnings$height - mean(earnings$height)
```

```
#Uniform
mod_unif <- stan_glm(earn ~ c_height, data=earnings, prior = NULL, prior_intercept = NULL, prior_aux = N

#Weak automatic priors
mod_weak <- stan_glm(earn ~ c_height, data=earnings, refresh=0)

#Informative priors
mod_inform <- stan_glm(earn ~ c_height, prior=normal(0,1000), prior_intercept = normal(20000,2500), data

mod_inform_2 <- stan_glm(earn ~ c_height, prior = normal(0,250), prior_intercept = normal(20000, 2500),
```

```
## [1] "=================UNIFORM PRIOR===================="

## stan_glm
##  family:       gaussian [identity]
##  formula:      earn ~ c_height
##  observations: 1816
##  predictors:   2
## ------
##             Median  MAD_SD
## (Intercept) 21140.2   514.4
## c_height     1590.9   135.0
##
## Auxiliary parameter(s):
##       Median  MAD_SD
## sigma 21699.1   366.5
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg

## [1] "===================WEAK PRIOR====================="

## stan_glm
##  family:       gaussian [identity]
##  formula:      earn ~ c_height
##  observations: 1816
##  predictors:   2
## ------
##             Median  MAD_SD
## (Intercept) 21148.5   501.7
## c_height     1593.2   135.7
##
## Auxiliary parameter(s):
##       Median  MAD_SD
## sigma 21695.4   350.7
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg

## [1] "===================INFORMATIVE PRIOR===================="
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      earn ~ c_height
##  observations: 1816
##  predictors:   2
## ------
##             Median  MAD_SD
## (Intercept) 21095.5   489.5
## c_height     1570.6   131.7
##
## Auxiliary parameter(s):
##       Median  MAD_SD
## sigma 21700.9   365.4
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

```
## [1] "===================INFORMATIVE (BUT BAD) PRIOR===================="
```

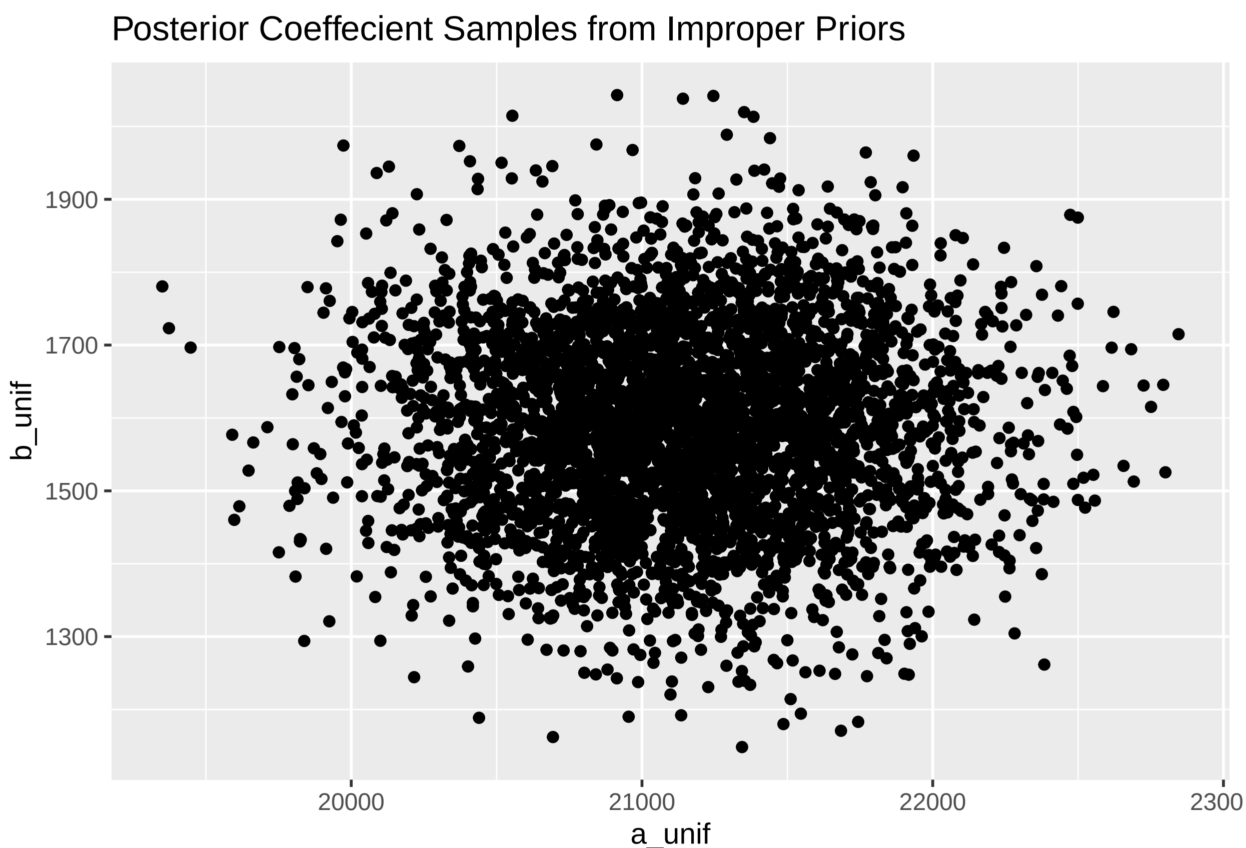
```
## stan_glm
##  family:       gaussian [identity]
##  formula:      earn ~ c_height
##  observations: 1816
##  predictors:   2
## ------
##             Median  MAD_SD
## (Intercept) 21111.9   476.6
## c_height     1242.3   121.9
##
## Auxiliary parameter(s):
##       Median  MAD_SD
## sigma 21742.9   364.4
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

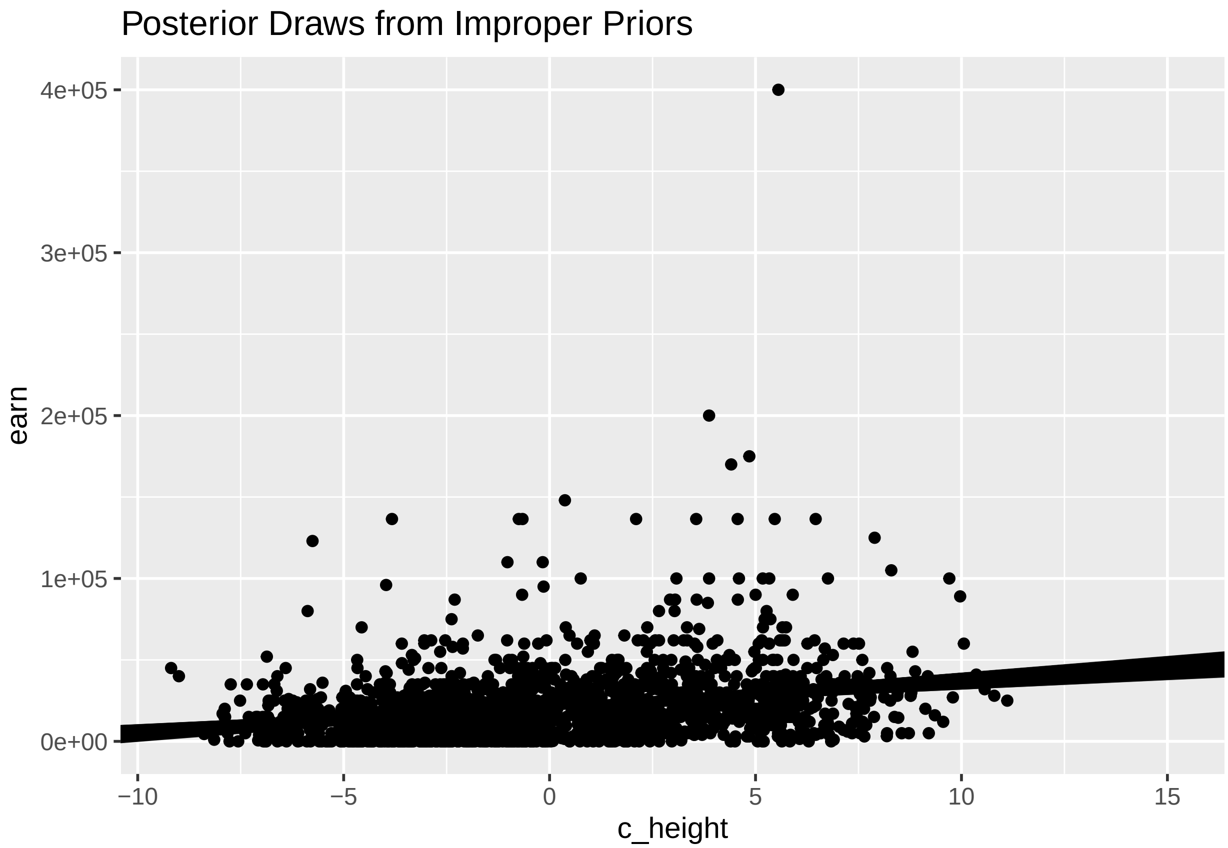
We see that the inferences in this case are essentially equivalent independent of out choice of prior except in the case of the final model with too restrictive of a prior. This is because the data are more informative than any of the first three priors and so affect the posterior more than the priors do. In the case of the last prior we are telling the model that we have to have really strong evidence in the data to conculde that the relationship between height and earnings to be outside of -250 to 250 dollars/inch.

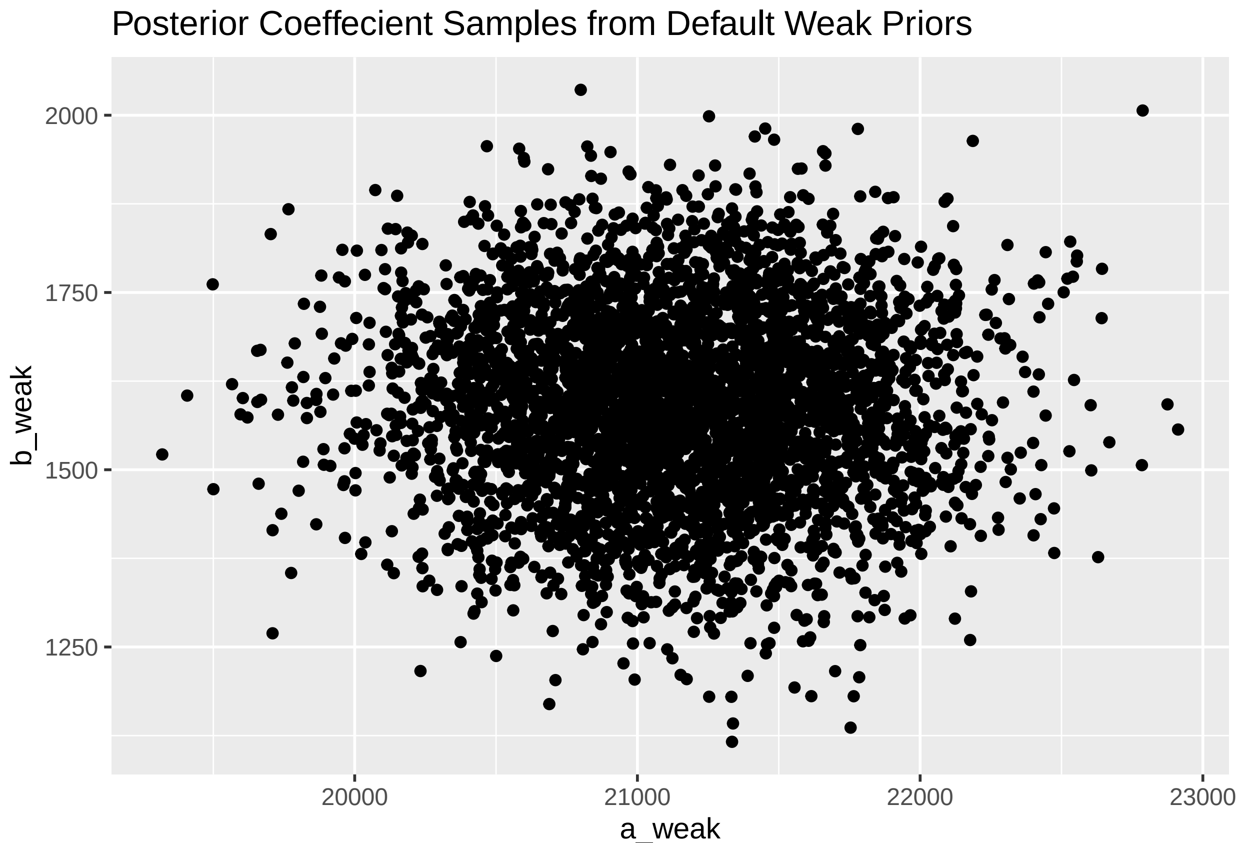
We will perform posterior predictive checks for each model to visualize these results.

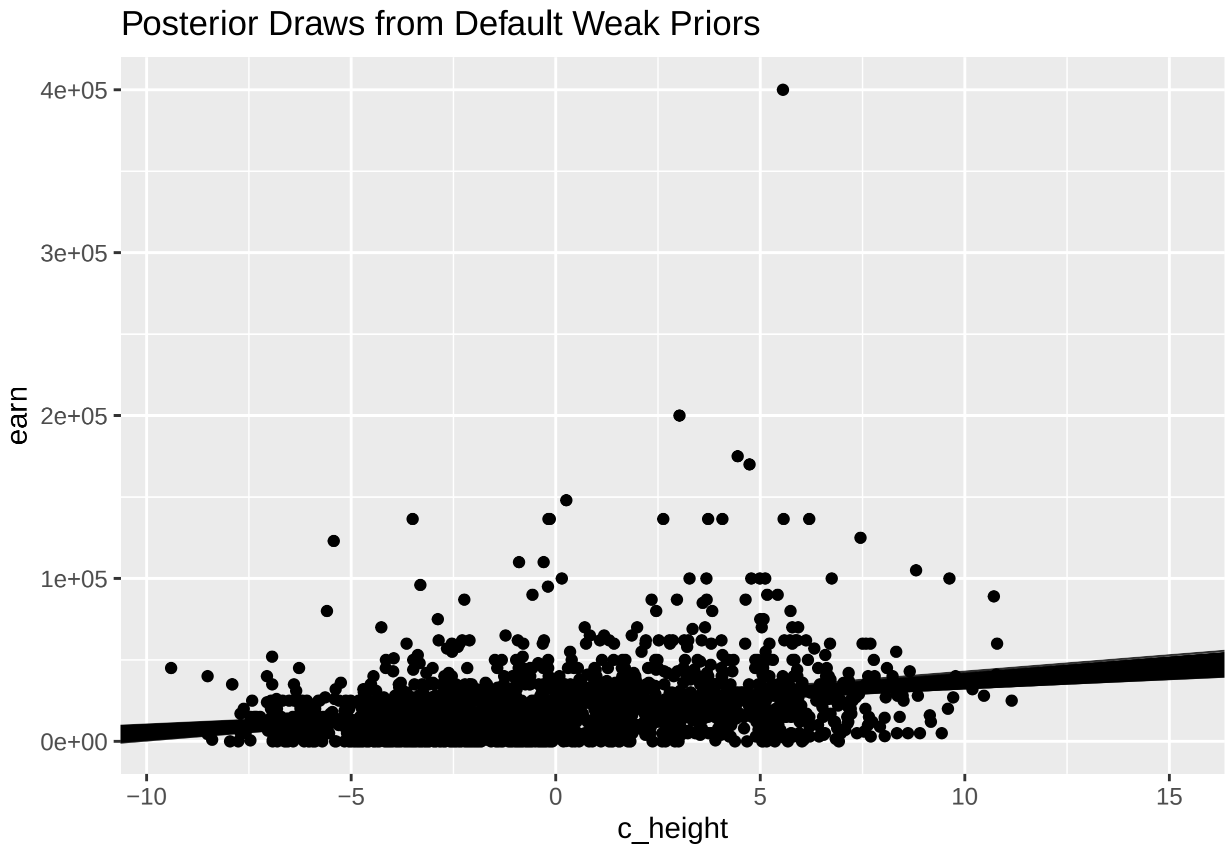
```
#Posterior Draws

sims_unif <- as.matrix(mod_unif)
a_unif <- sims_unif[,1]
b_unif <- sims_unif[,2]
unif <- data.frame(a_unif, b_unif)
```

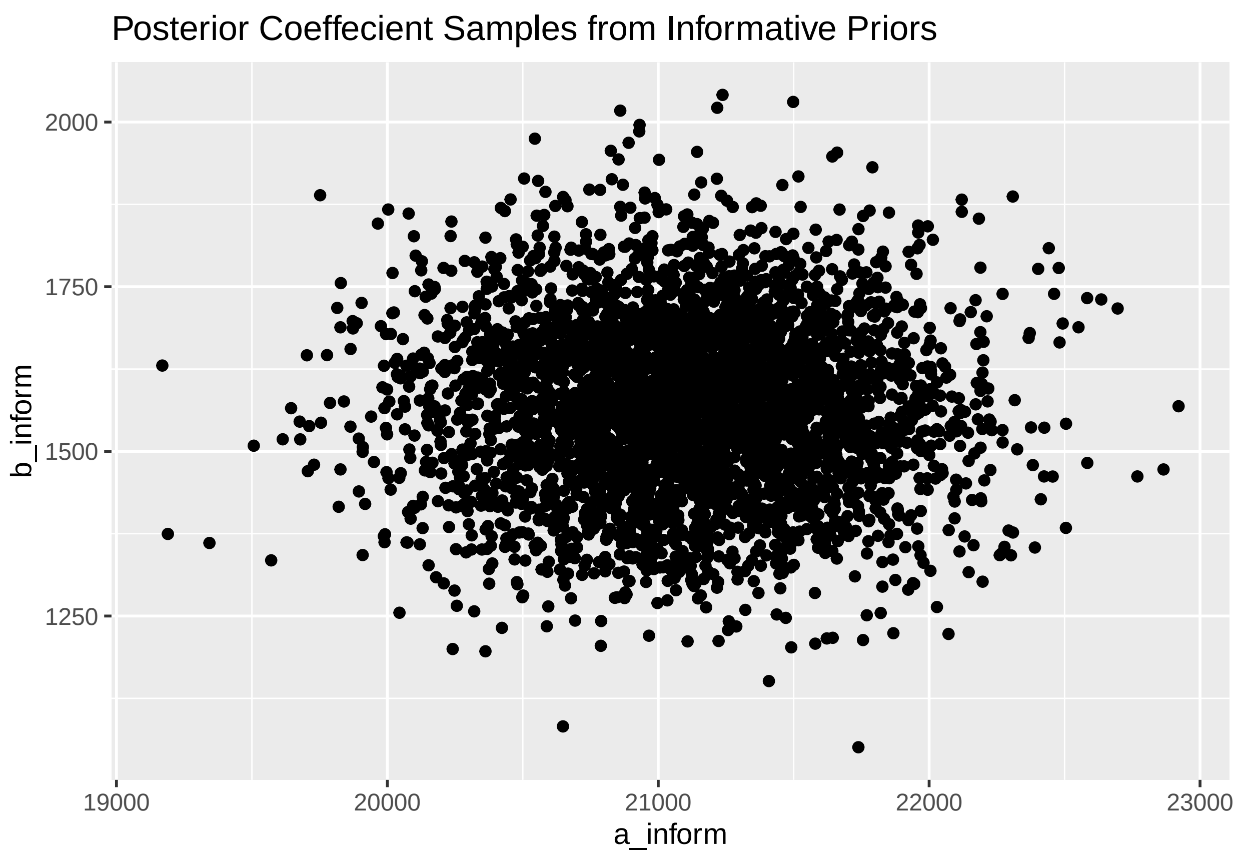
```r
sims_weak <- as.matrix(mod_weak)
a_weak <- sims_weak[,1]
b_weak <- sims_weak[,2]
weak <- data.frame(a_weak, b_weak)

sims_inform <- as.matrix(mod_inform)
a_inform <- sims_inform[,1]
b_inform <- sims_inform[,2]
inform <- data.frame(a_inform, b_inform)

sims_inform_2 <- as.matrix(mod_inform_2)
a_inform_2 <- sims_inform_2[,1]
b_inform_2 <- sims_inform_2[,2]
inform_2 <- data.frame(a_inform_2, b_inform_2)
```



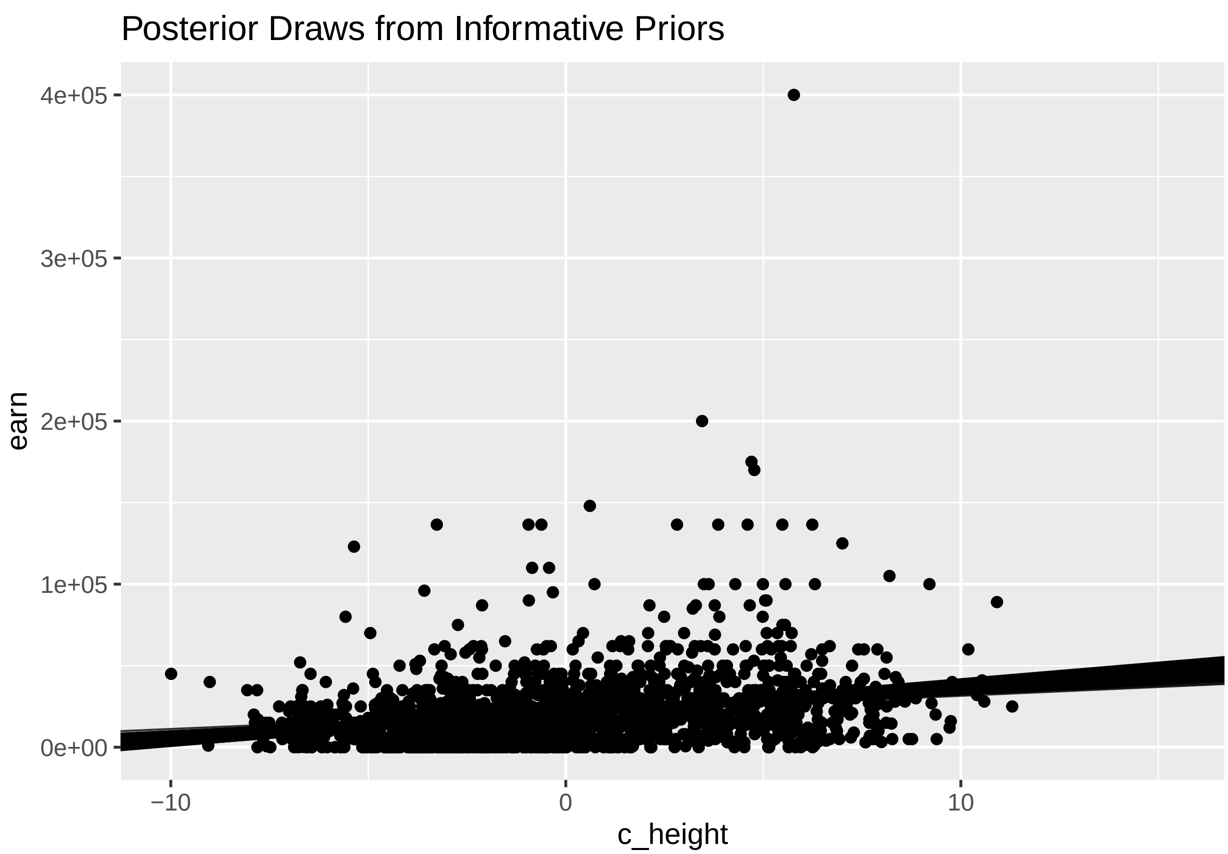

Posterior Draws from Improper Priors
4e+05
3e+05
2e+05
1e+05
0e+00
earn
−10
−5
0
5
10
15
c_height

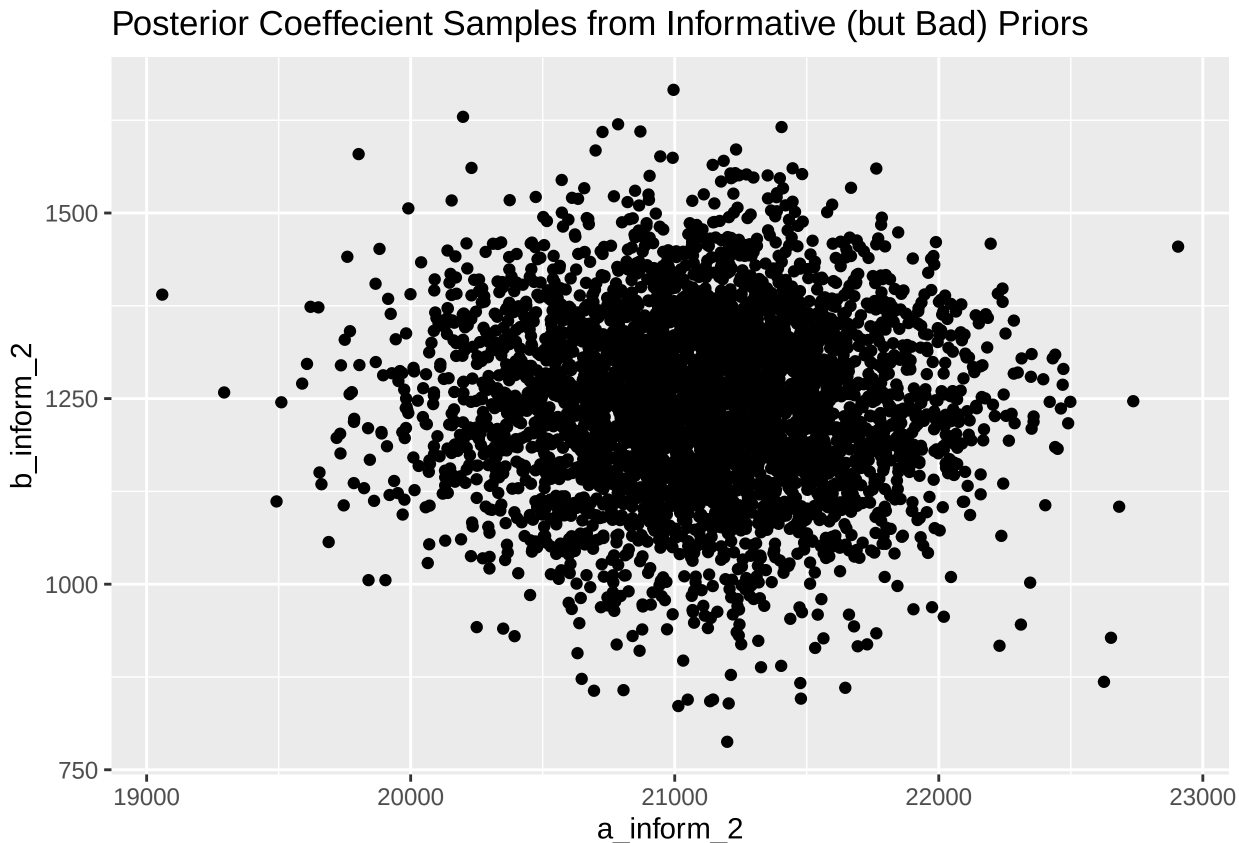

Posterior Coeffecient Samples from Default Weak Priors
2000
1750
1500
1250
b_weak
20000
21000
22000
23000
a_weak


Posterior Draws from Default Weak Priors
4e+05
3e+05
2e+05
1e+05
0e+00
earn
−10
−5
0
5
10
15
c_height

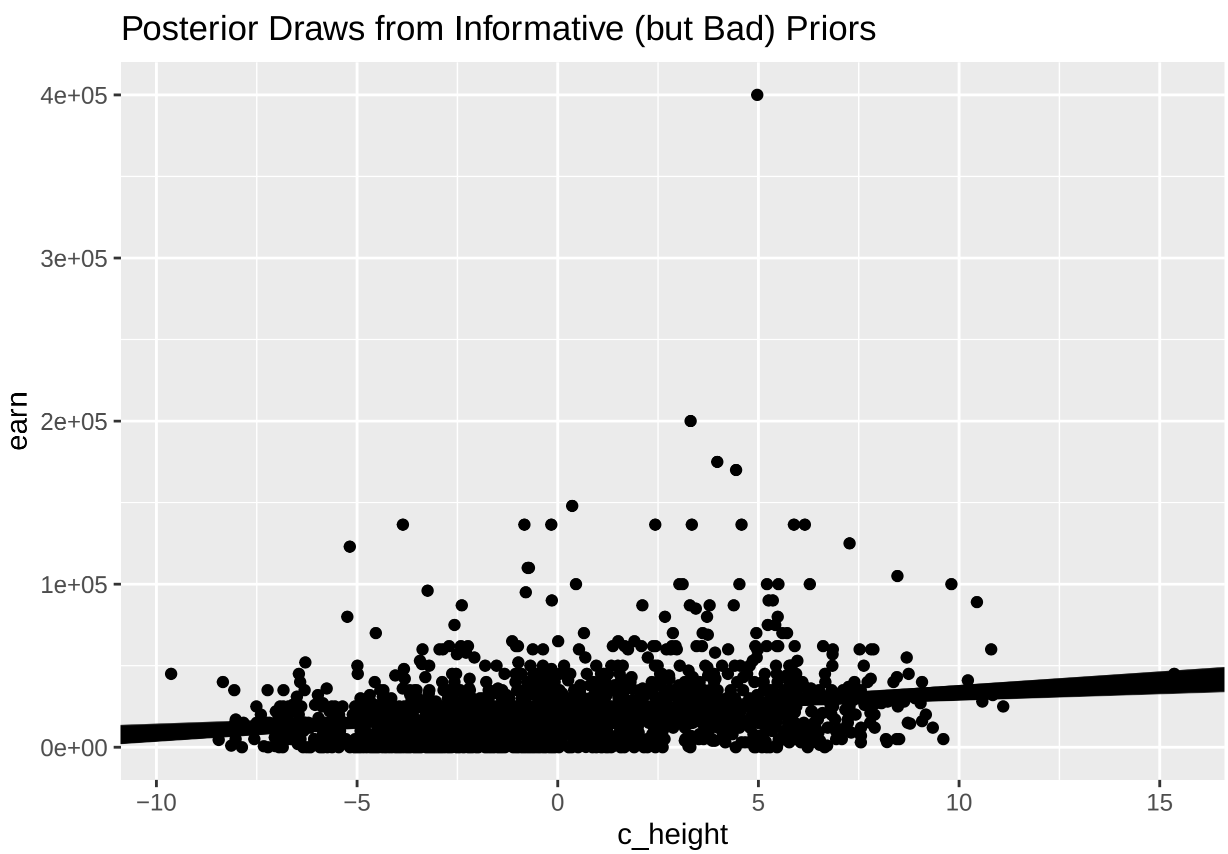


Posterior Coeffecient Samples from Informative Priors
2000
1750
1500
1250
b_inform
19000
20000
21000
22000
23000
a_inform


Posterior Draws from Informative Priors
earn
4e+05
3e+05
2e+05
1e+05
0e+00
−10
0
10
c_height



Posterior Coeffecient Samples from Informative (but Bad) Priors
b_inform_2
1500
1250
1000
750
19000
20000
21000
22000
23000
a_inform_2

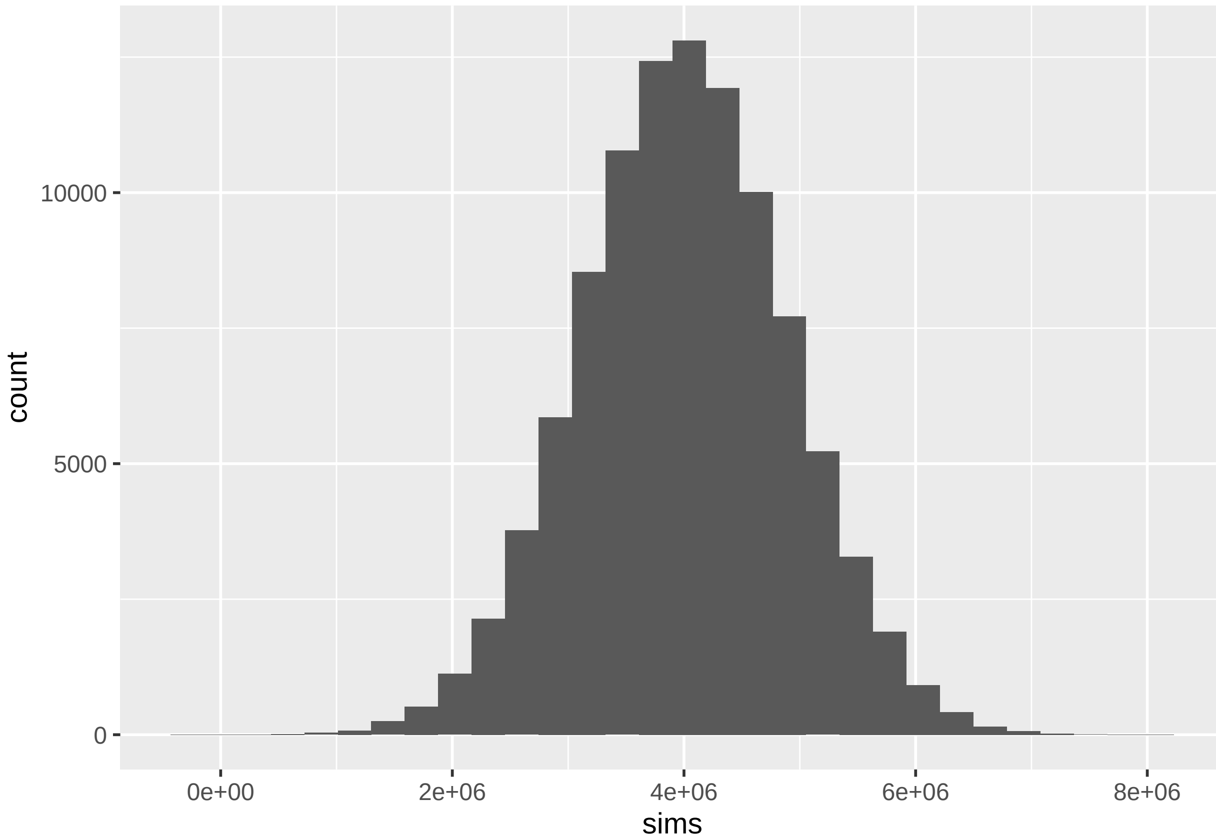


Question 8

We will assume that the resulting sales increase from each minute of advertisement is independent from each other minute of advertisement.

```
sims <- c()

for (i in 1:100000){
  sim <- rnorm(20,500000,200000) - 300000

  sims <- c(sims, sum(sim))

}

ggplot() + aes(sims) + geom_histogram()
```

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




```
## [1] "Simulated expected return on ad campaign:, 4002708.77445752"

## [1] "Simulated probability of negative net gain: 3e-05"
```

Question 9

Here is a model with reasonable priors for the slope and intercept based on the information stated in the problem.

$$y \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta(x - \bar{x})$$

$$\alpha \sim \text{Normal}(75, 10)$$

$$\beta \text{ Normal}(0.65, 0.15)$$