

## 9.2)

a) Estimates of the posterior betas

|       | npreg       | bp        | skin       | bmi       | ped       | age       |
|-------|-------------|-----------|------------|-----------|-----------|-----------|
| 2.5%  | -1.92632167 | 0.4158359 | -0.2502817 | 0.7957794 | 6.143916  | 0.6205851 |
| 97.5% | -0.03556841 | 0.7867952 | 0.3956118  | 1.7380682 | 20.831784 | 1.2573272 |

b) Estimates of the after best model selection

|       | npreg      | bp        | skin      | bmi       | ped      | age       |
|-------|------------|-----------|-----------|-----------|----------|-----------|
| 2.5%  | -0.9547727 | 0.0000000 | 0.0000000 | 0.4947543 | 0.000000 | 0.4790979 |
| 97.5% | 0.0000000  | 0.3199292 | 0.3256697 | 1.3485814 | 17.21936 | 0.9904431 |

Probability of best model

|    | all         | prob  |
|----|-------------|-------|
| 1  | 0 0 0 1 1 1 | 0.465 |
| 2  | 0 0 0 1 0 1 | 0.228 |
| 3  | 0 1 0 1 1 1 | 0.104 |
| 4  | 1 0 0 1 1 1 | 0.045 |
| 5  | 0 0 1 1 1 1 | 0.035 |
| 6  | 1 0 0 1 0 1 | 0.029 |
| 7  | 0 1 0 1 0 1 | 0.025 |
| 8  | 0 0 1 1 0 1 | 0.019 |
| 9  | 1 1 0 1 1 1 | 0.014 |
| 10 | 0 1 1 1 1 1 | 0.010 |
| 11 | 0 0 1 0 1 1 | 0.007 |
| 12 | 1 1 0 1 0 1 | 0.004 |
| 13 | 0 1 1 0 1 1 | 0.003 |
| 14 | 1 0 1 1 1 1 | 0.003 |
| 15 | 0 1 1 1 0 1 | 0.002 |
| 16 | 1 1 1 1 1 1 | 0.002 |
| 17 | 0 1 1 0 0 1 | 0.001 |
| 18 | 1 0 1 0 1 1 | 0.001 |
| 19 | 1 0 1 1 0 1 | 0.001 |
| 20 | 1 1 1 0 1 1 | 0.001 |
| 21 | 1 1 1 1 0 1 | 0.001 |

The 95% Confidence interval of the betas after model selection is much smaller than the initial posterior betas in the previous model

### 9.3)

a)

Posterior betas

|            |             |             |             |             |             |
|------------|-------------|-------------|-------------|-------------|-------------|
| M          | So          | Ed          | Po1         | Po2         | LF          |
| 0.27678836 | 0.00524724  | 0.54078190  | 1.43765701  | -0.76412976 | -0.06027487 |
| M.F        | Pop         | NW          | U1          | U2          | GDP         |
| 0.12848061 | -0.06790958 | 0.09862836  | -0.26956973 | 0.37235594  | 0.22350019  |
| Ineq       | Prob        | Time        |             |             |             |
| 0.71217333 | -0.27432825 | -0.05484386 |             |             |             |

95% Confidence Posterior Betas

|       |            |            |            |             |            |            |
|-------|------------|------------|------------|-------------|------------|------------|
|       | M          | So         | Ed         | Po1         | Po2        | LF         |
| 2.5%  | 0.03291574 | -0.3144854 | 0.2103544  | -0.08949473 | -2.4374830 | -0.3506106 |
| 97.5% | 0.52363143 | 0.3213659  | 0.8555000  | 3.02151656  | 0.8014371  | 0.1864203  |
|       | M.F        | Pop        | NW         | U1          | U2         | GDP        |
| 2.5%  | -0.1540167 | -0.2827156 | -0.2172904 | -0.63424502 | 0.06664789 | -0.2287955 |
| 97.5% | 0.3968335  | 0.1579139  | 0.4045775  | 0.06595465  | 0.70201893 | 0.6709754  |
|       | Ineq       | Prob       | Time       |             |            |            |
| 2.5%  | 0.2560089  | -0.5117502 | -0.2912905 |             |            |            |
| 97.5% | 1.1211167  | -0.0280633 | 0.1821423  |             |            |            |

LS Coefficients

|               |               |               |               |               |
|---------------|---------------|---------------|---------------|---------------|
| (Intercept)   | M             | So            | Ed            | Po1           |
| -0.0004581088 | 0.2865181425  | -0.0001140461 | 0.5445140840  | 1.4716210675  |
| Po2           | LF            | M.F           | Pop           | NW            |
| -0.7817801428 | -0.0659645619 | 0.1312980228  | -0.0702919266 | 0.1090566856  |
| U1            | U2            | GDP           | Ineq          | Prob          |
| -0.2705364468 | 0.3687303043  | 0.2380594756  | 0.7262919898  | -0.2852264319 |
| Time          |               |               |               |               |
| -0.0615768625 |               |               |               |               |

LS 95% Confidence Interval

|        |             |             |             |            |            |             |
|--------|-------------|-------------|-------------|------------|------------|-------------|
|        | (Intercept) | M           | So          | Ed         | Po1        | Po2         |
| 2.5 %  | -0.1614436  | 0.009556068 | -0.3754927  | 0.1781959  | -0.1939343 | -2.5186164  |
| 97.5 % | 0.1605274   | 0.563480217 | 0.3752646   | 0.9108322  | 3.1371764  | 0.9550562   |
|        | LF          | M.F         | Pop         | NW         | U1         | U2          |
| 2.5 %  | -0.3789230  | -0.1851921  | -0.3291440  | -0.2415737 | -0.6715677 | 0.001906366 |
| 97.5 % | 0.2469939   | 0.4477881   | 0.1885602   | 0.4596871  | 0.1304948  | 0.735554243 |
|        | GDP         | Ineq        | Prob        | Time       |            |             |
| 2.5 %  | -0.2900794  | 0.2487395   | -0.55809545 | -0.3288022 |            |             |
| 97.5 % | 0.7661984   | 1.2038444   | -0.01235741 | 0.2056485  |            |             |

The coefficients of the LS and Bayesian models are very similar and both agree on which variables are significant. M, Ed, U2, Ineq, Prob are the significant variables in the model, because their 95% confidence interval does not cross zero. M, Ed, U2, and Ineq all contribute to the crime rate as they increase and Prob decreases the crime rate as it increases.

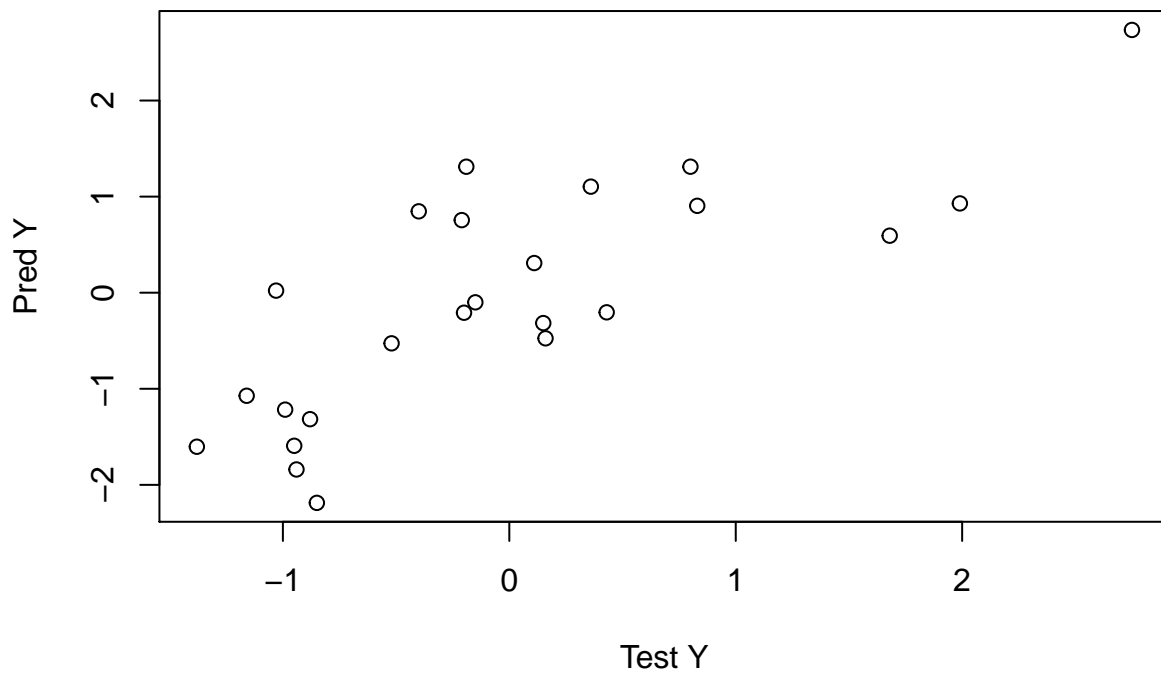
b)

In this particular case the linear regression error is less than the bayesian error. The bayesian estimates are not the same as the least squared estimates and the influence of the prior prevents them from equaling the least square estimates. In the case when the random sample of data is not representative of the population, the least squares estimates will hurt prediction accuracy on the test set. The Bayesian estimates are more consistent so in the long run it has a lower average error.

Least Squares Regression Error

```
[1] 0.5562224
```

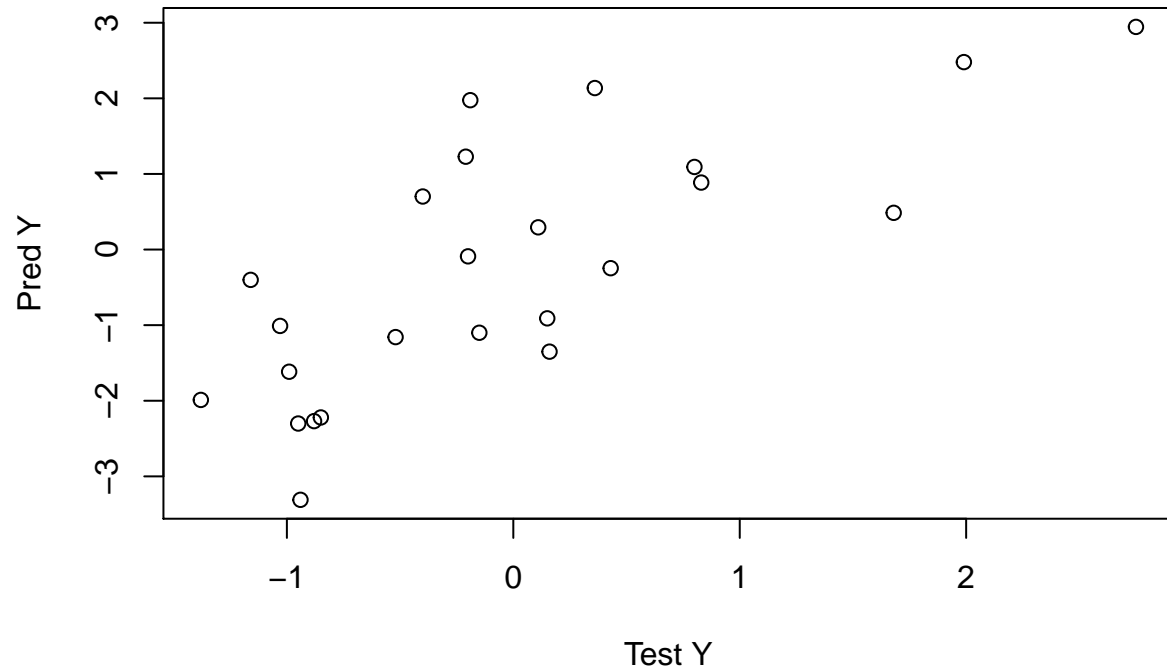
### LS Regression



Bayesian Regression Error

[1] 1.280432

## Bayesian Regression



c)

I ran 1000 simulations of the experiment and computed the prediction error for the least squares and bayesian regression. On average the bayesian regression has a smaller mean and variance error than the LS regression.

| ls.error        | bs.error       |
|-----------------|----------------|
| Min. : 0.3265   | Min. :0.3297   |
| 1st Qu.: 0.7566 | 1st Qu.:0.6806 |
| Median : 1.0086 | Median :0.9519 |
| Mean : 1.2169   | Mean :1.1382   |
| 3rd Qu.: 1.4099 | 3rd Qu.:1.3218 |
| Max. :15.0421   | Max. :8.0762   |

Variance of estimates

|          | ls.error    | bs.error    |
|----------|-------------|-------------|
| ls.error | 0.753775305 | 0.008040009 |
| bs.error | 0.008040009 | 0.549300280 |

