Course Capstone Project

Lytemar

October 9, 2017

Introduction

The data set that we will analyze is the opening response time (ms) of an electromechanical valve. The valve controls the flow of several fluids where the opening and closing is actuated by a solenoid. The shutoff seal is an elastomer that is attached to the end of the solenoid plunger. In this project, we will determine how the seal vintage, fluid and applied coil voltage affects the valve's opening response time.

Exploratory Data Analysis

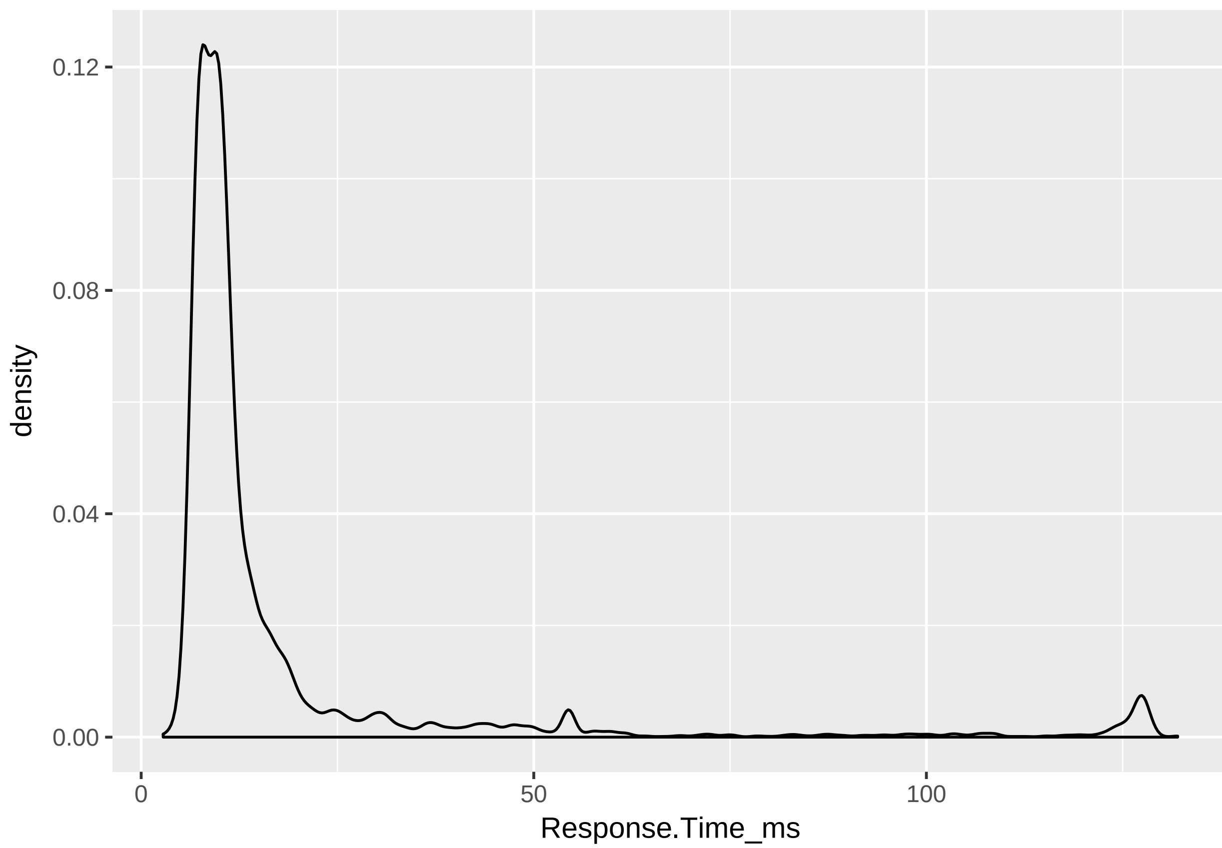
The summary of the data set is

```
##  Valve.Rev  Fluid    Time.Between.Actuations_hr Plunger.Age Coil.Age
##  New:2082   A: 845   Min.   :  0.0              New:2692    New:2722
##  Old:3118   B: 400   1st Qu.:  0.0              Old:2508    Old:2478
##             C: 378   Median : 24.0
##             D: 479   Mean   : 32.5
##             E:1059   3rd Qu.: 24.0
##             F:2039   Max.   :336.0
##  Voltage   Seal.Type Seal.Age   Response.Time_ms
##  24:1099   A:5080    New:2922   Min.   :  2.80
##  30:4101   B: 120    Old:2278   1st Qu.:  8.00
##                                 Median : 10.00
##                                 Mean   : 17.64
##                                 3rd Qu.: 14.40
##                                 Max.   :132.00
```

where *Valve.Rev* indicates the revision of the valve design, *Fluid* indicates the different fluids as masked factors, *Time.Between.Actuations_hr* indicates the time between valve actuations in hours, **.Age* indicates the age of the indicated component, *voltage* is the applied DC solenoid coil voltage, *Seal.Type* indicates the types of seal materials as masked factors, and *Response.Time_ms* is the time to open the valve in ms.

Plots

In the interest of report length, we only show the following density plot that illustrates the distribution of response time. Additionally, box plots of response time versus the other variables would show how response time varies due to the other variables in the data set. Here, we see that the distribution of response time is right-skewed, therefore we will apply a log transformation before fitting a Bayesian regression model.



Modeling

Here we will fit a Bayesian linear model from which we will see which predictors affect response time. The coefficients of the predictors are

- b[1]: intercept,
- b[2]: is.new.vlv,
- b[3]: Fluid,
- b[4]: Time.Between.Actuations_ms,
- b[5]: is.new.plunger,
- b[6]: is.new.coil,
- b[7]: is.24v,
- b[8]: is.seal_A,
- b[9]: is.new.seal,
- b[10]: Fluid*is.24v,
- b[11]: is.new.seal*is.24v, and
- b[12]: is.new.vlv*is.24v.

Finally, we will compute the probability that the response time is less than 100 ms given Voltage from our MCMC results.

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 5200
```

```
##    Unobserved stochastic nodes: 13
##    Total graph size: 46982
##
## Initializing model
```

Using normal likelihood, with normal priors on the coefficients and inverse gamma prior for the variance, with 5000 burn-in iterations, and 100000 total iterations for the 3 chains, we get the following autocorrelation diagnostics

```
##             b[1]        b[2]       b[3]        b[4]      b[5]        b[6]
## Lag 0  1.0000000  1.00000000 1.00000000 1.000000000 1.0000000 1.000000000
## Lag 1  0.9897978  0.60664432 0.89018821 0.372614287 0.9670880 0.844716605
## Lag 5  0.9543573  0.13056208 0.61553393 0.034468609 0.8542509 0.416826607
## Lag 10 0.9141722  0.02882825 0.40901049 0.017671493 0.7463280 0.169074108
## Lag 50 0.6328113 -0.01162318 0.01120933 0.002514952 0.3642070 0.004286718
##              b[7]      b[8]      b[9]      b[10]       b[11]        b[12]
## Lag 0  1.00000000 1.0000000 1.0000000 1.00000000 1.000000000  1.000000000
## Lag 1  0.87346294 0.9882002 0.9649568 0.83749209 0.472367675  0.426955360
## Lag 5  0.54859280 0.9541707 0.8718205 0.49392680 0.105016707  0.034094529
## Lag 10 0.34494075 0.9175593 0.7857858 0.31283091 0.033637630  0.007495166
## Lag 50 0.01494924 0.6314727 0.4169199 0.01146919 0.004418294 -0.001102957
##                sig2
## Lag 0   1.000000000
## Lag 1   0.001181056
## Lag 5   0.002337366
## Lag 10  0.001957570
## Lag 50 -0.002790058
```

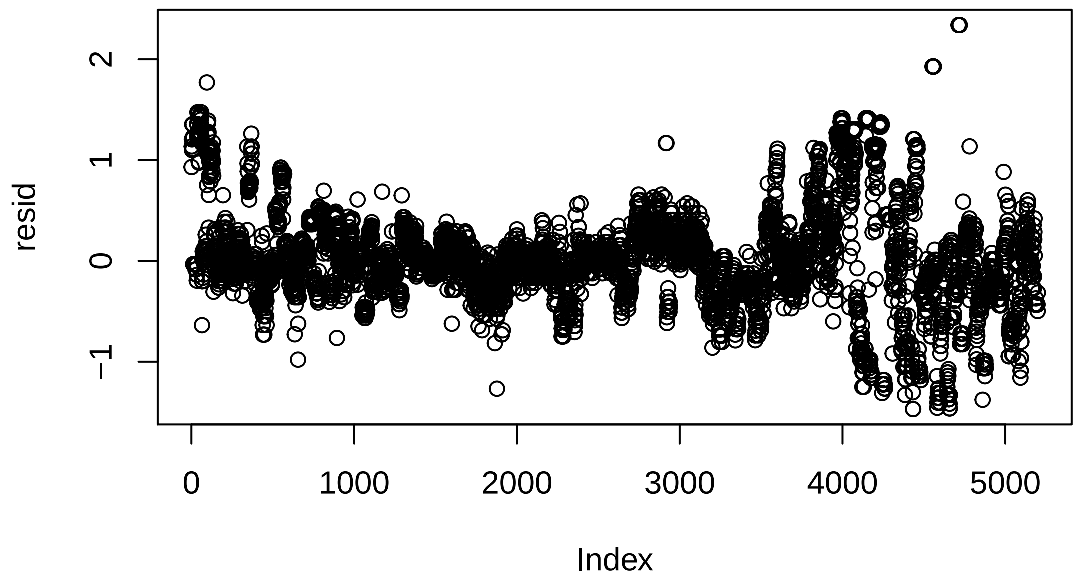
Here, we see that not all of the parameters have converged, giving the following effective sample sizes

```
##       b[1]       b[2]       b[3]       b[4]       b[5]       b[6]
##   1302.264  62278.894  12792.504 105622.969   3090.199  26381.478
##       b[7]       b[8]       b[9]      b[10]      b[11]      b[12]
##  14162.338   1423.093   2721.551  15250.927  72293.169 107064.372
##       sig2
## 296289.279
```

However, we are interested in the means of the coefficients which are

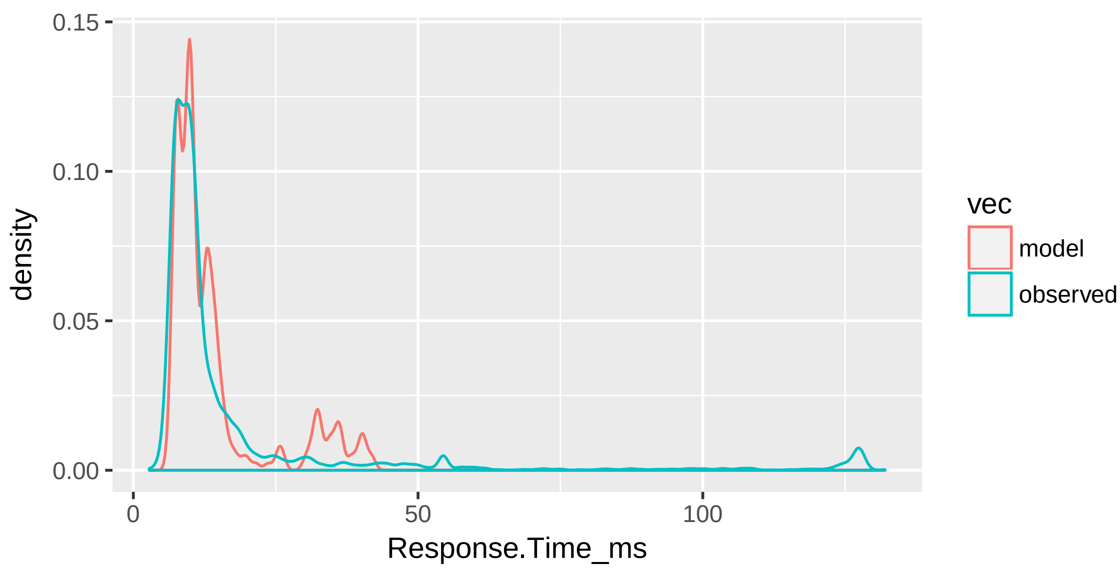
```
##         b[1]         b[2]         b[3]         b[4]         b[5]
##  2.468528122 -0.288401295 -0.027887898  0.001506414  0.323745397
##         b[6]         b[7]         b[8]         b[9]        b[10]
## -0.132448535  1.070648737  0.212210981 -0.443149433 -0.025315172
##        b[11]        b[12]         sig2
## -0.185497724 -0.432900517  0.231887618
```

The plot of the residuals shows that we have some dependency on the variance with respect to the data point. This would need to be remedied by a different transformation on the data.



Results

The following plot superinposes the log response time distributions of the data and resulting model.



Finally, using the modeled posterior distribution, the probability that the response time given 30 volts is less than given 24 volts.

```
## [1] 0.8839307
```