Week09 Homework

Brian Ritz

Saturday, March 07, 2015

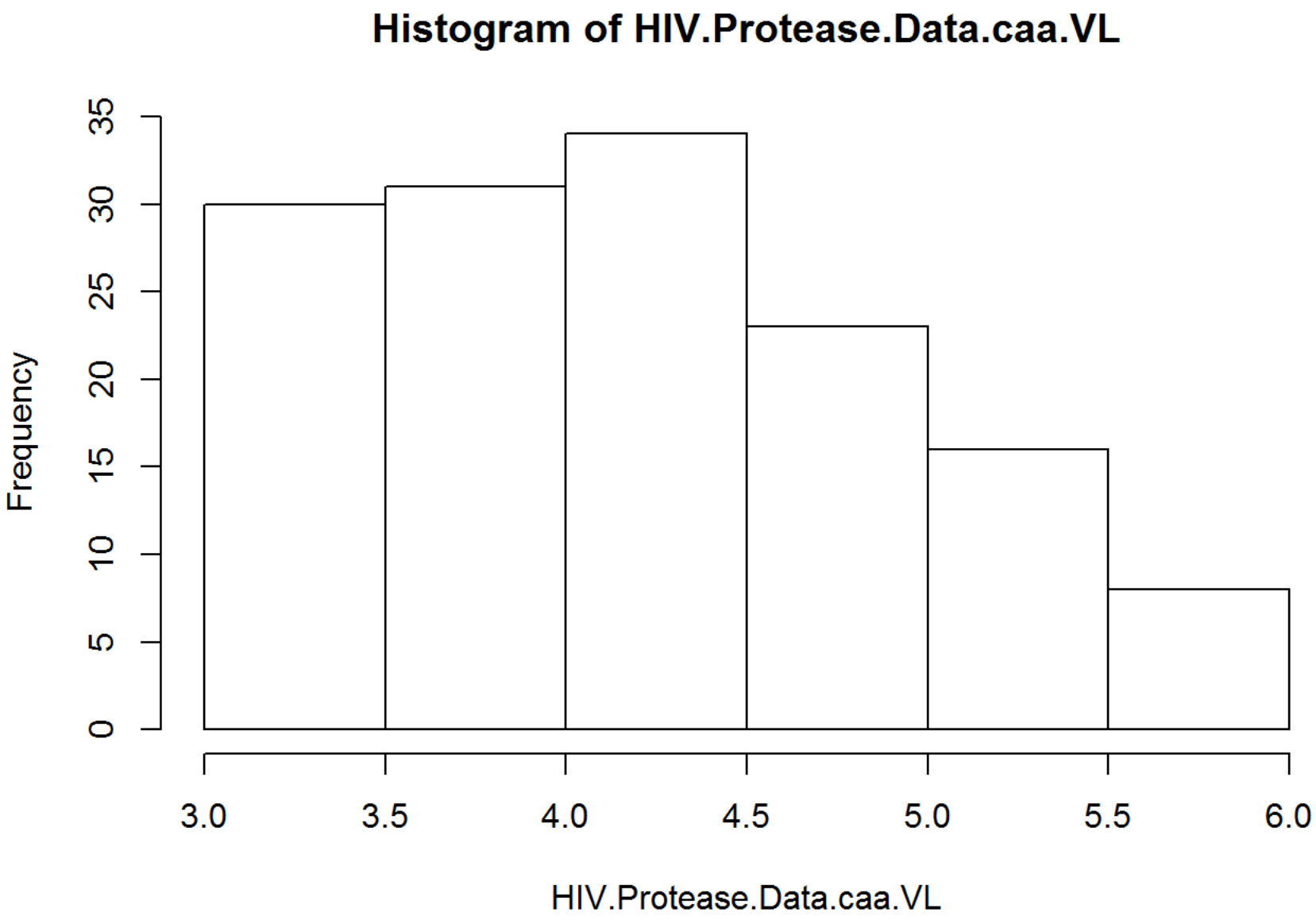
```
HIV.Protease.Data<-read.csv(file="HIV_Protease.csv",header=TRUE,sep=",")
HIV.Protease.Data<-as.data.frame(HIV.Protease.Data)
head(HIV.Protease.Data)
```

```
##   Resp  VL CODON_2
## 1    0 4.3     caa
## 2    0 3.6     caa
## 3    0 3.2     caa
## 4    0 5.7     caa
## 5    0 3.5     caa
## 6    0 3.9     caa
```

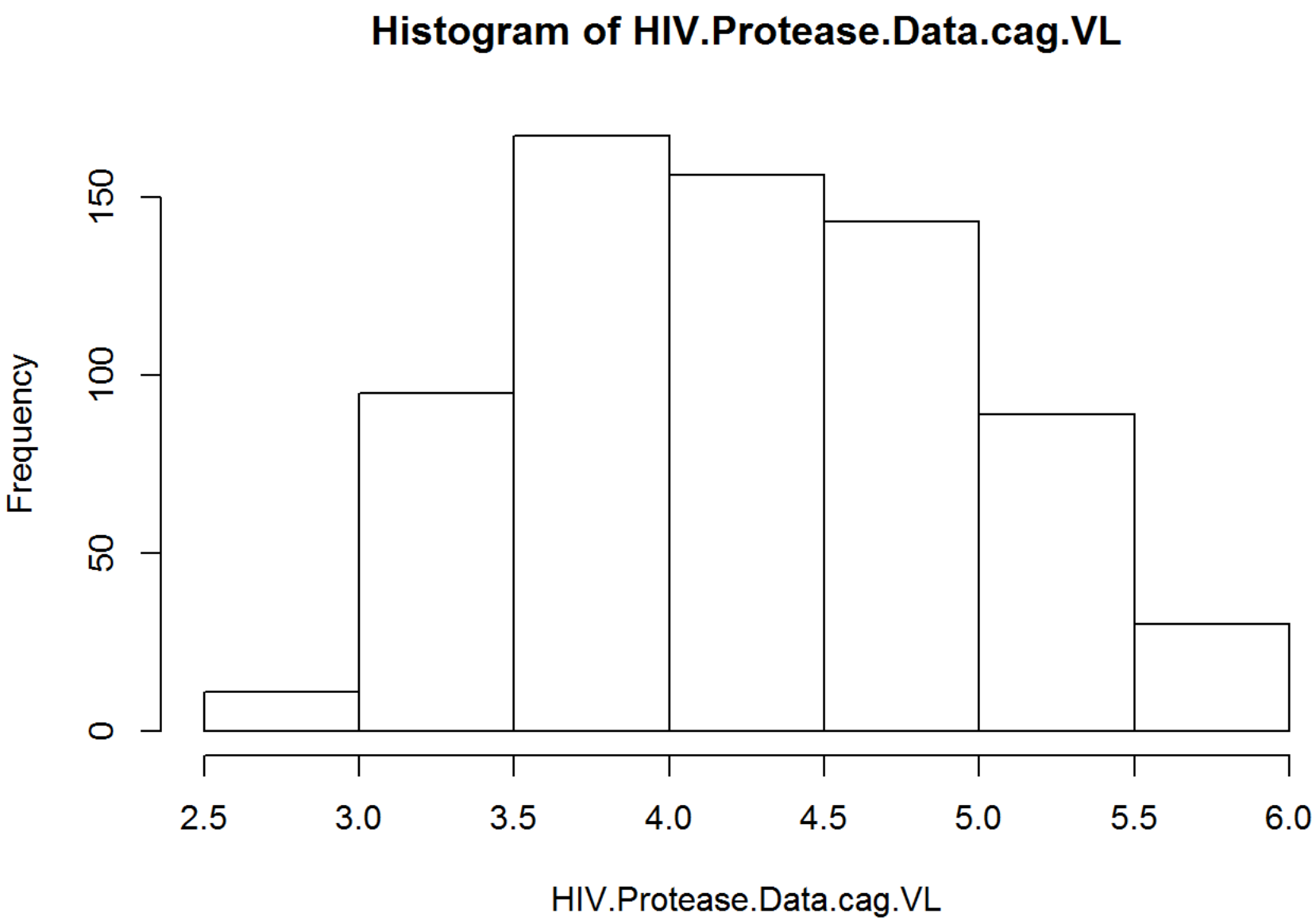
```
table(HIV.Protease.Data$CODON_2)
```

```
##
## caa cag cak car maa
## 142 691   1  67   1
```

Show a histogram of the viral loads of the Dataset when CODON_2 = caa:



Show a histogram of the viral loads of the Dataset when CODON_2 = cag:



We will run a model to determine if a change from caa to cag will play a significant role in the failure of the drug.

First we compare the means:

```
c(mean(HIV.Protease.Data.cag.VL),mean(HIV.Protease.Data.caa.VL))
```

```
## [1] 4.314616 4.248239
```

The means look similar, but we will need to run statistical tests to confirm that they are not the same.

Fit a linear model:

```
lm.CODON_2.VL <- lm(VL~CODON_2, data=HIV.Protease.Data)
AIC(lm.CODON_2.VL)
```

```
## [1] 1931.876
```

```
anova(lm.CODON_2.VL)
```

```
## Analysis of Variance Table
##
## Response: VL
##            Df Sum Sq Mean Sq F value Pr(>F)
## CODON_2     4   2.44 0.60906  1.2312 0.2959
## Residuals 897 443.72 0.49468
```

From the anova results we can see that the means do not appear different. This is because the p-value on the CODON_2 response is greater than our threshhold of .05. This is an indicator that perhaps CODON_2 does not make a difference in the viral load.

Now we will fit the model with a random effect.

```
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 3.1.3
```

```
## Loading required package: Matrix
## Loading required package: Rcpp
```

```
lm.CODON_2.VL.Mixed <- lmer(VL ~ 1 + (1|CODON_2), data=HIV.Protease.Data)
AIC(lm.CODON_2.VL.Mixed)
```

```
## [1] 1936.485
```

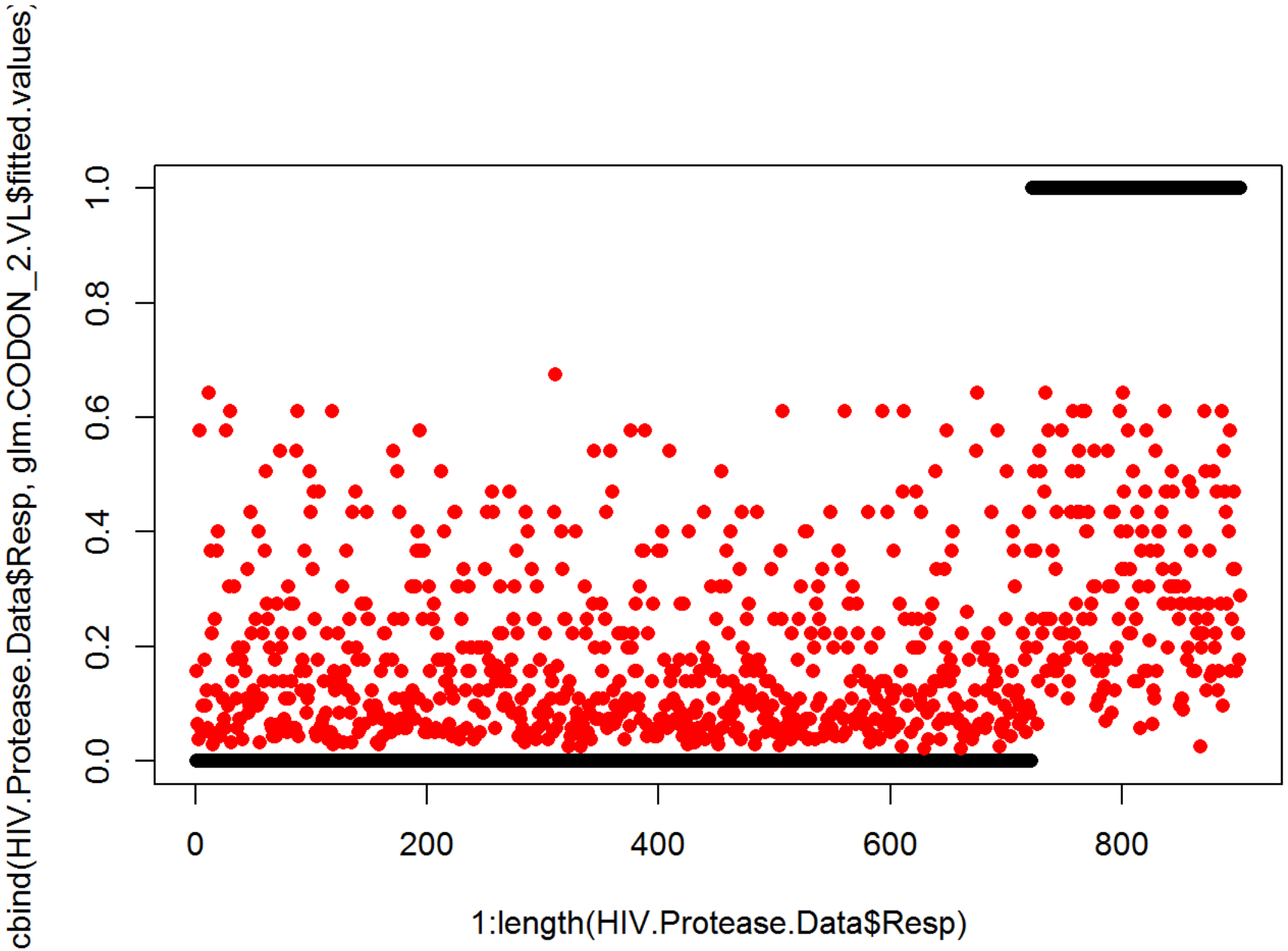
The AIC for the random effect is greater than for the fixed effects model. This is surprising in the context of this assignement, but I conclude that the linear model is a better fit to the data.

Now we will fit a logistic regression:

```
glm.CODON_2.VL <- glm(Resp~VL, family="binomial", data=HIV.Protease.Data)
print(AIC(glm.CODON_2.VL))
```

```
## [1] 782.068
```

```
matplot(1:length(HIV.Protease.Data$Resp),cbind(HIV.Protease.Data$Resp,glm.CODON_2.VL$fitted.values),pch=16)
```




```
head(cbind(HIV.Protease.Data$Resp,glm.CODON_2.VL$fitted.values))
```

```
##   [,1]       [,2]
```

```
## 1    0 0.15714185
## 2    0 0.06457383
## 3    0 0.03765410
## 4    0 0.57625950
## 5    0 0.05651215
## 6    0 0.09557494
```

We will fit a model with CODON_2 as a predictor:

```
library(MASS)

# glm.CODON_2.LV <- nlme(Resp~VL+CODON_2, data=HIV.Protease.Data)
glm.CODON_2.LV2 <- glmmPQL(Resp~VL, random=~1|CODON_2, family=binomial, data=HIV.Protease.Data)
```

```
## Loading required package: nlme
##
## Attaching package: 'nlme'
##
## The following object is masked from 'package:lme4':
##
##     lmList
##
## iteration 1
## iteration 2
## iteration 3
## iteration 4
## iteration 5
```

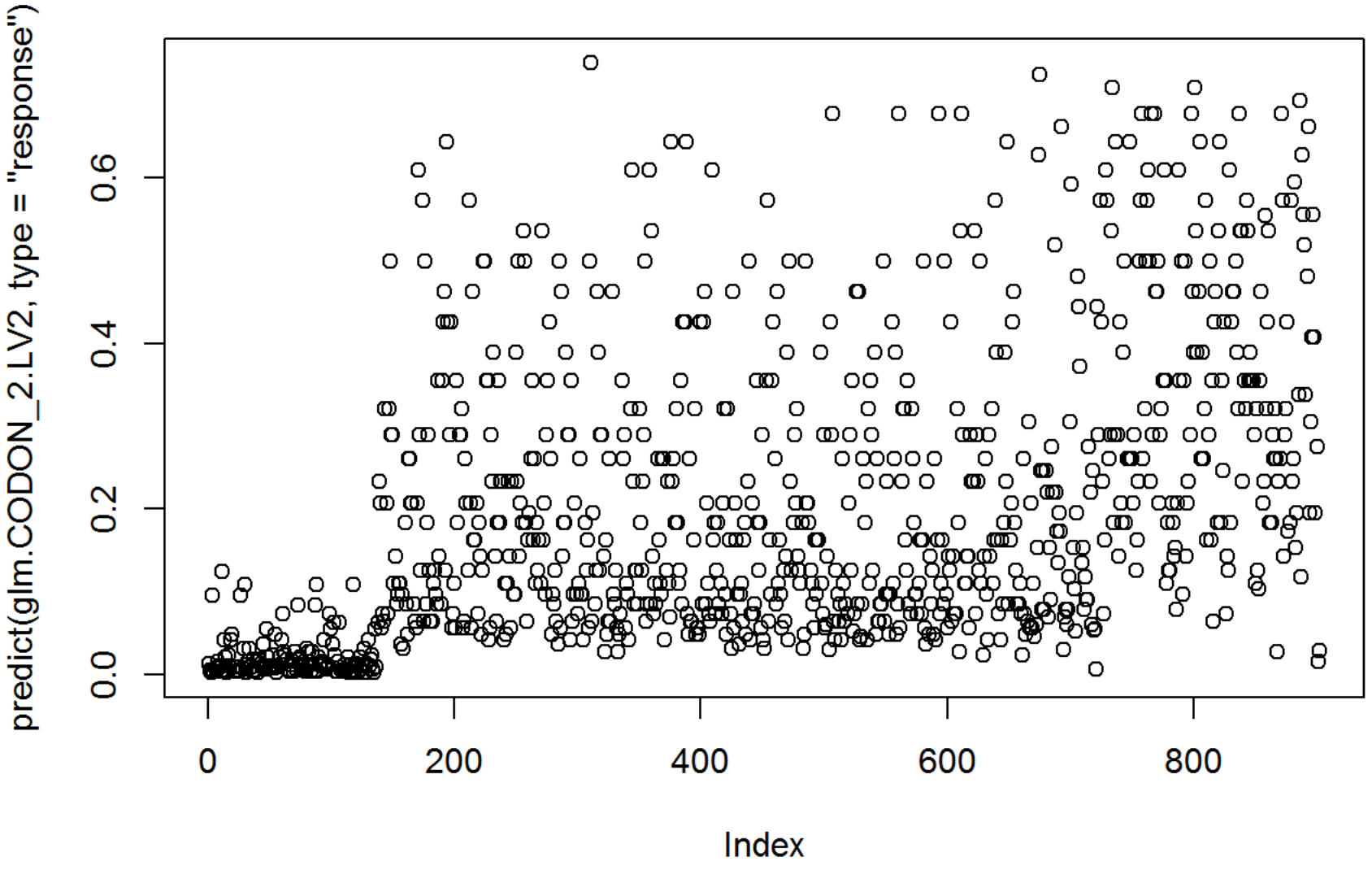
```
# glm.CODON_2.LV <- glm(Resp~VL+CODON_2, family = binomial, data=HIV.Protease.Data)
```

Compare the results to other models:

```
AIC(glm.CODON_2.LV2)
```

```
## [1] NA
```

```
plot(predict(glm.CODON_2.LV2,type="response"))
```



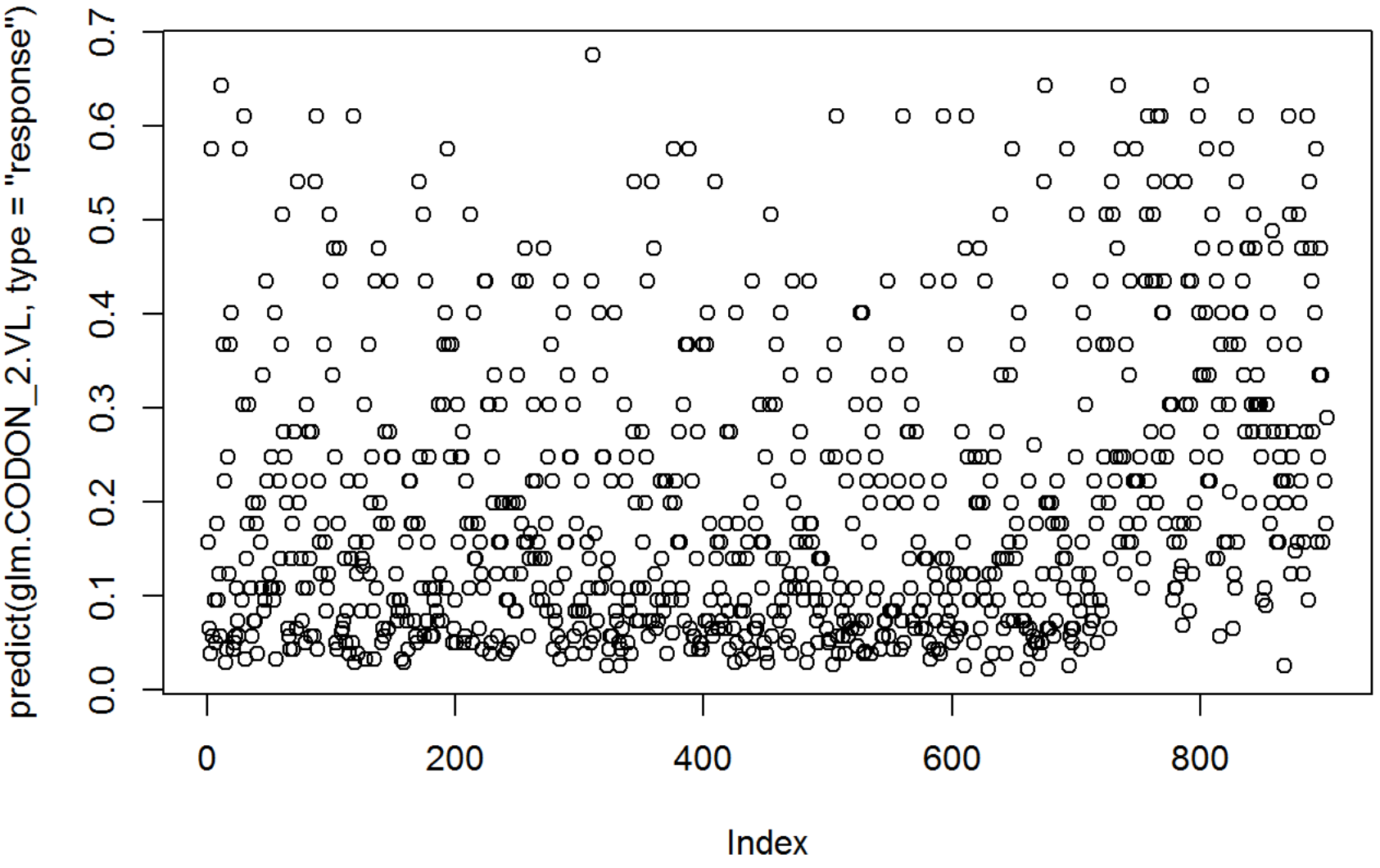

```
head(cbind(HIV.Protease.Data$Resp,predict(glm.CODON_2.LV2,type="response")))
```

```
##      [,1]        [,2]
```

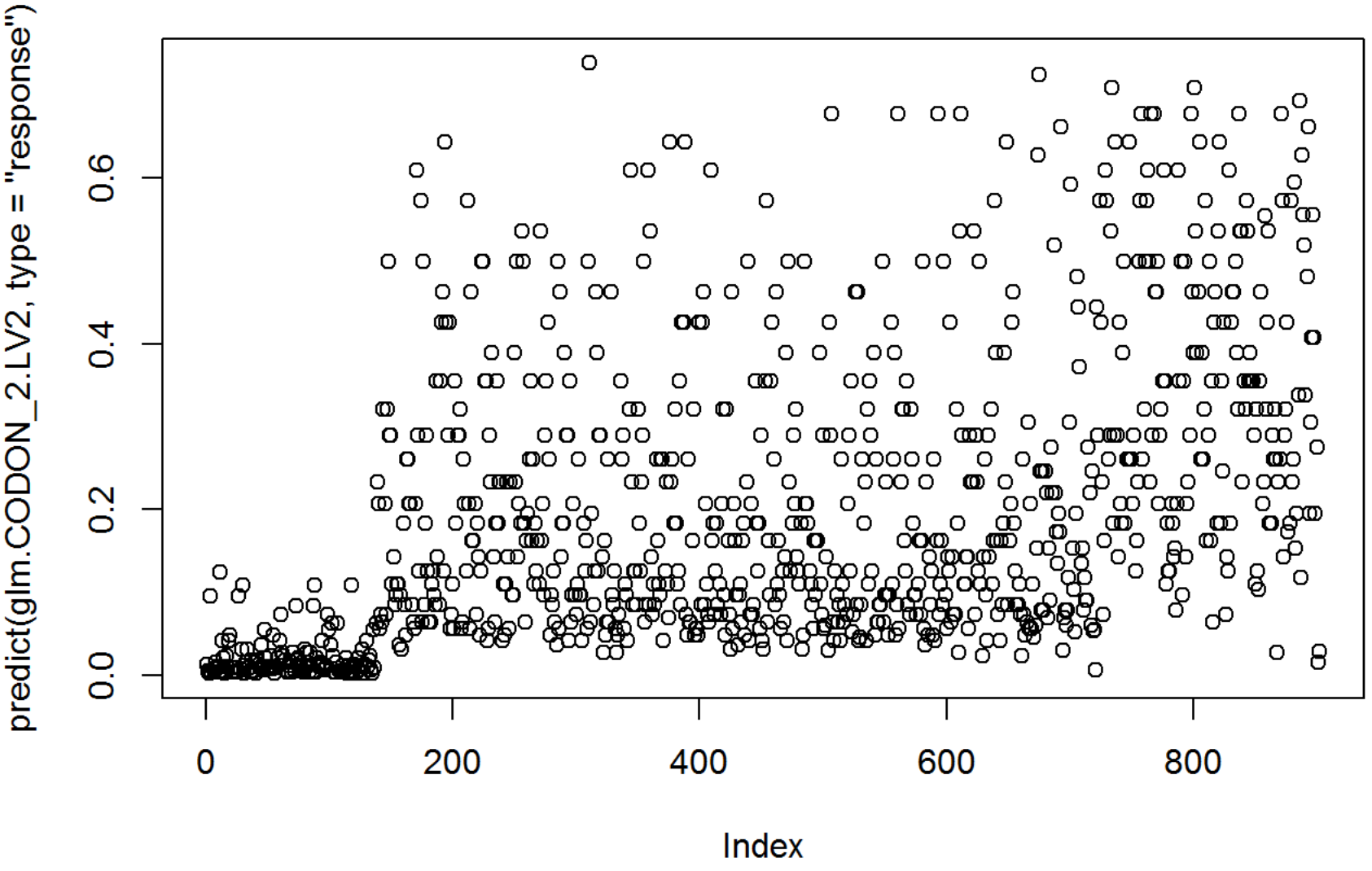
```
## caa    0 0.012893407
## caa    0 0.004582969
## caa    0 0.002530831
## caa    0 0.095129182
## caa    0 0.003951190
## caa    0 0.007146782
```

Compare the predictions:

```
par(mfrow=c(1,1))
plot(predict(glm.CODON_2.VL, type="response"))
```




```
plot(predict(glm.CODON_2.LV2, type="response"))
```



The plots of the predictions do look different. The model with the random effect of the CODON_2 variable has a concentration of predicitons near zero, which the regular logistic regression does not have.