# Biomial regression

## Chandan Kumar Pandey

## 2022-10-17

## Biomial regression modeling

In the Binomial linear regression, we model for following equation

$$log(p/(1-p)) = \alpha + \beta i * Xi$$

Solve for p

$$p = 1/1 + e^{-(\alpha+\beta i * Xi)}$$

## With single predictor

**Data**

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
bone_fract <- read.csv("./data/bmd.csv",header = T)
summary(bone_fract)
```

```
##       id              age             sex              fracture
##  Min.   :   35    Min.   :35.81   Length:169         Length:169
##  1st Qu.: 2018    1st Qu.:54.42   Class :character   Class :character
##  Median : 6702    Median :63.49   Mode  :character   Mode  :character
##  Mean   : 9103    Mean   :63.63
##  3rd Qu.:17100    3rd Qu.:72.08
##  Max.   :24208    Max.   :88.75
##    weight_kg        height_cm      medication         waiting_time
##  Min.   :36.00    Min.   :142.0   Length:169         Min.   : 5.00
##  1st Qu.:56.00    1st Qu.:154.0   Class :character   1st Qu.: 9.00
##  Median :64.50    Median :160.5   Mode  :character   Median :14.00
##  Mean   :64.67    Mean   :160.3                      Mean   :19.74
##  3rd Qu.:73.00    3rd Qu.:166.0                      3rd Qu.:24.00
##  Max.   :96.00    Max.   :177.0                      Max.   :96.00
##       bmd
##  Min.   :0.4076
##  1st Qu.:0.6708
##  Median :0.7861
##  Mean   :0.7831
##  3rd Qu.:0.8888
##  Max.   :1.3624
```

```r
boxplot(bmd~fracture,data = bone_fract)
```
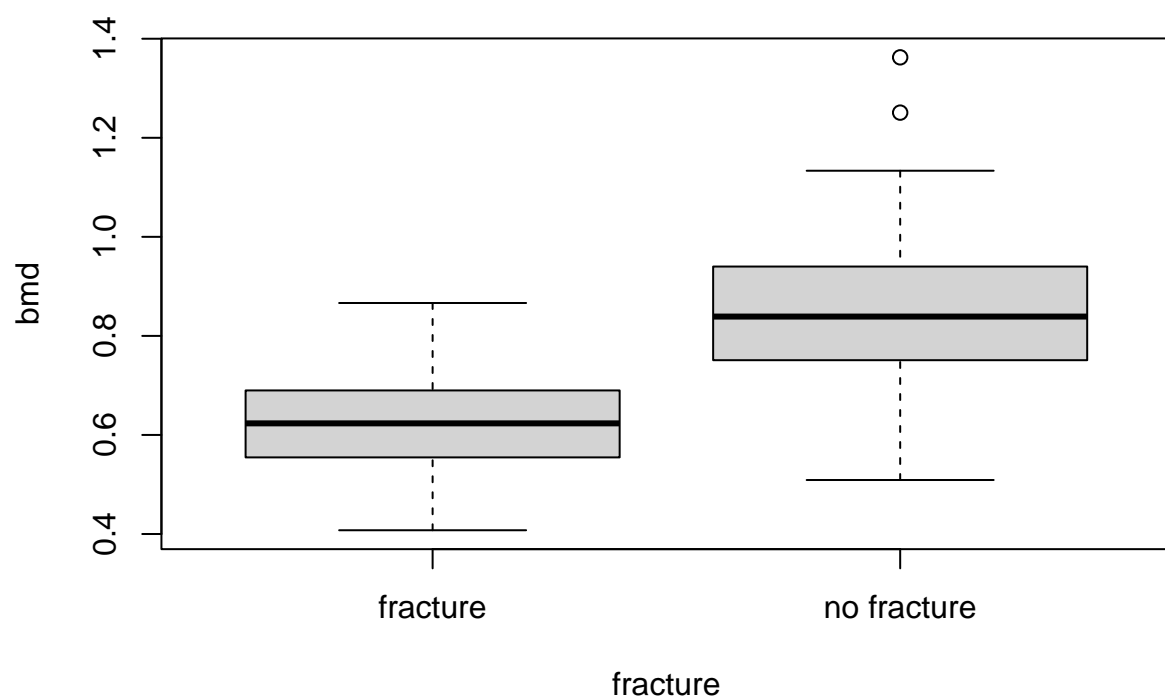
```
bone_fract$fracture <- ifelse(bone_fract$fracture=="fracture",1,0)

ggpairs(bone_fract[,c(4,9)])
```
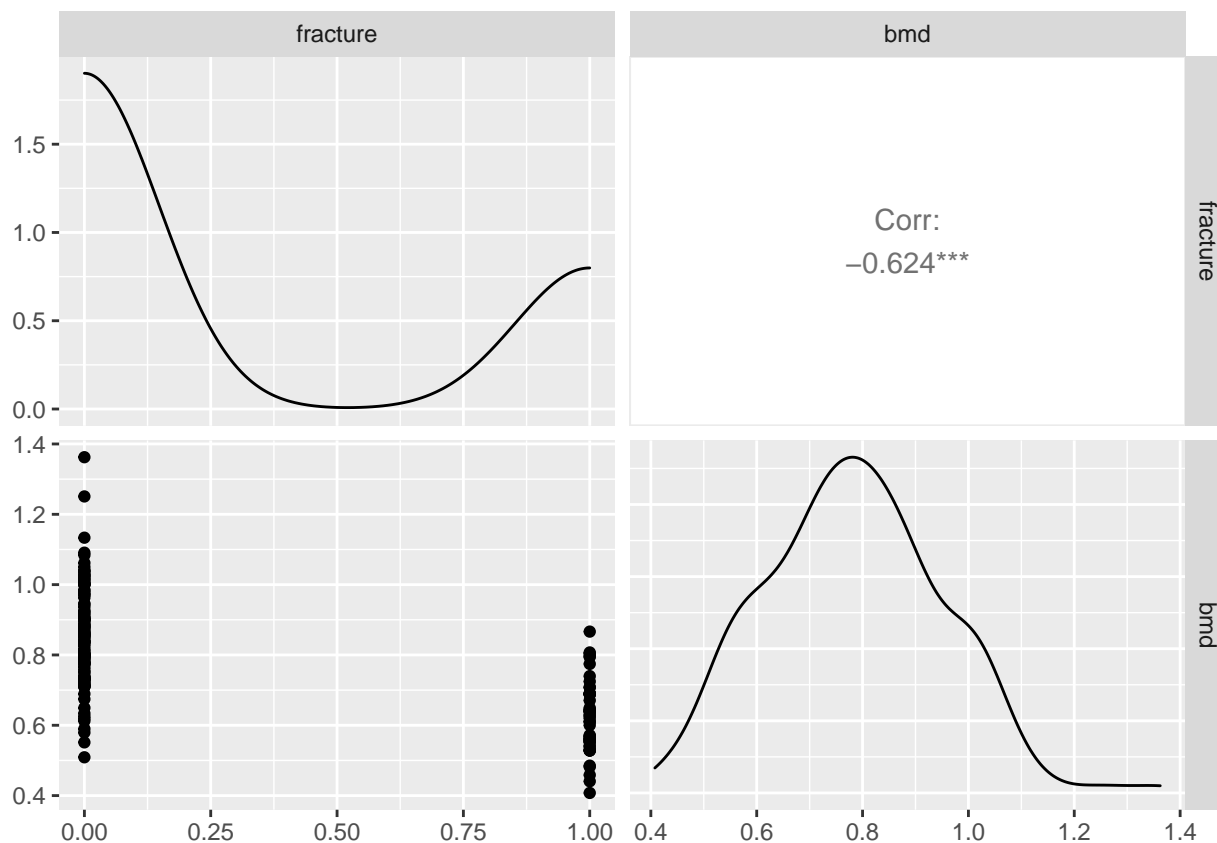
```
mod1 <- glm(fracture~bmd,data = bone_fract,family = "binomial")
summary(mod1)
```

```
##
## Call:
## glm(formula = fracture ~ bmd, family = "binomial", data = bone_fract)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3003  -0.5191  -0.2083   0.4453   2.4960
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   10.607      1.767   6.002 1.95e-09 ***
## bmd          -15.785      2.497  -6.321 2.59e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 205.27  on 168  degrees of freedom
## Residual deviance: 116.02  on 167  degrees of freedom
## AIC: 120.02
##
## Number of Fisher Scoring iterations: 6
```

```
par(mfrow=c(2,2))
plot(mod1)
```
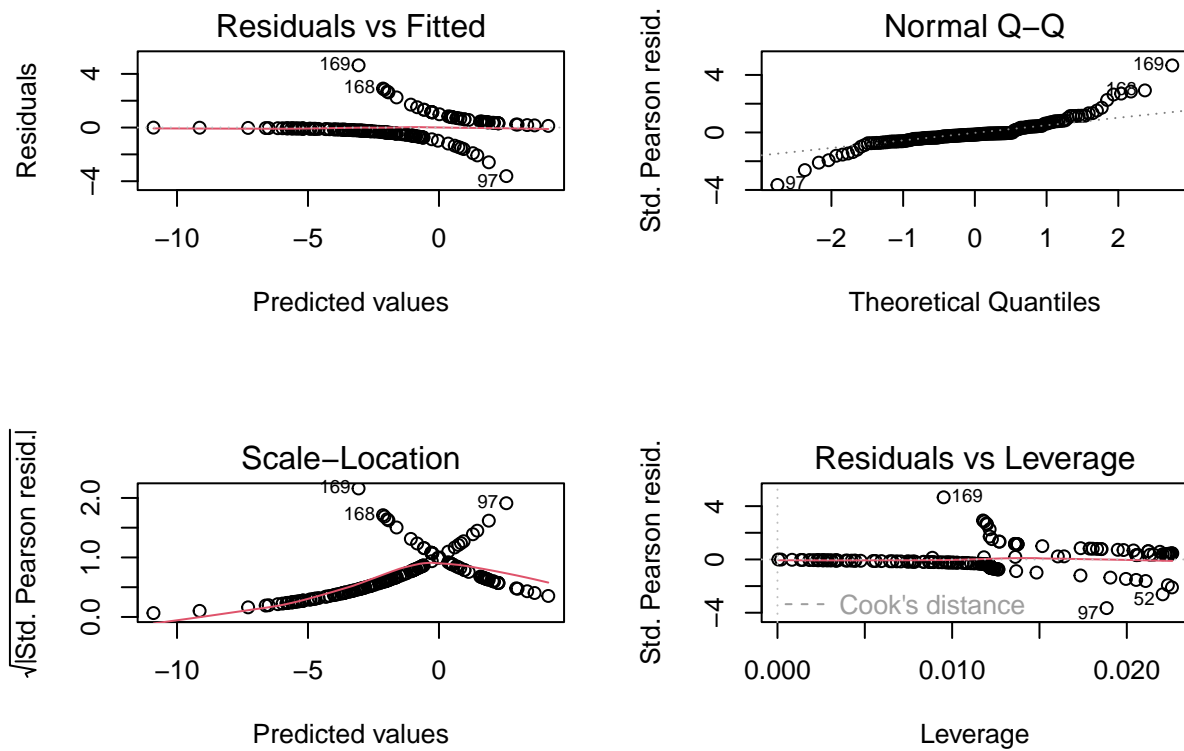


```
plot(mod1$residuals~mod1$fitted.values,col=as.factor(mod1$data$fracture))
PseudoR_value <- (mod1$null.deviance-mod1$deviance)/mod1$null.deviance
```
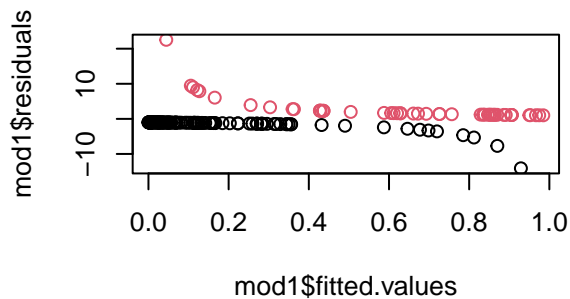
**Model is fit.**

We need to plot this model

```r
newdata <- data.frame(bmd=seq(min(bone_fract$bmd),max(bone_fract$bmd),by=0.01))
predicted_value <- predict(mod1,newdata,type = "response",se.fit = T)
predicted_value<-predicted_value%>%as.data.frame()
predicted_value<-cbind(predicted_value,bmd=newdata$bmd)
library(ggplot2)
plot_bmd <- ggplot(data = predicted_value,aes(x=bmd,y=fit))
plot_bmd+geom_line()+
  geom_ribbon(aes(ymax=fit+se.fit,ymin=fit-se.fit,fill="yellow"),alpha=0.1,
              show.legend = F)+
  geom_point(data = bone_fract,aes(x=bmd,y=fracture))+
  theme_classic()+
  labs(x="Bone mass density",y="Probabilty of fracture")+
  geom_text(aes(x=1.2,y=0.95,label=paste("Pseduo R^2 =", round(PseudoR_value,3))))
```
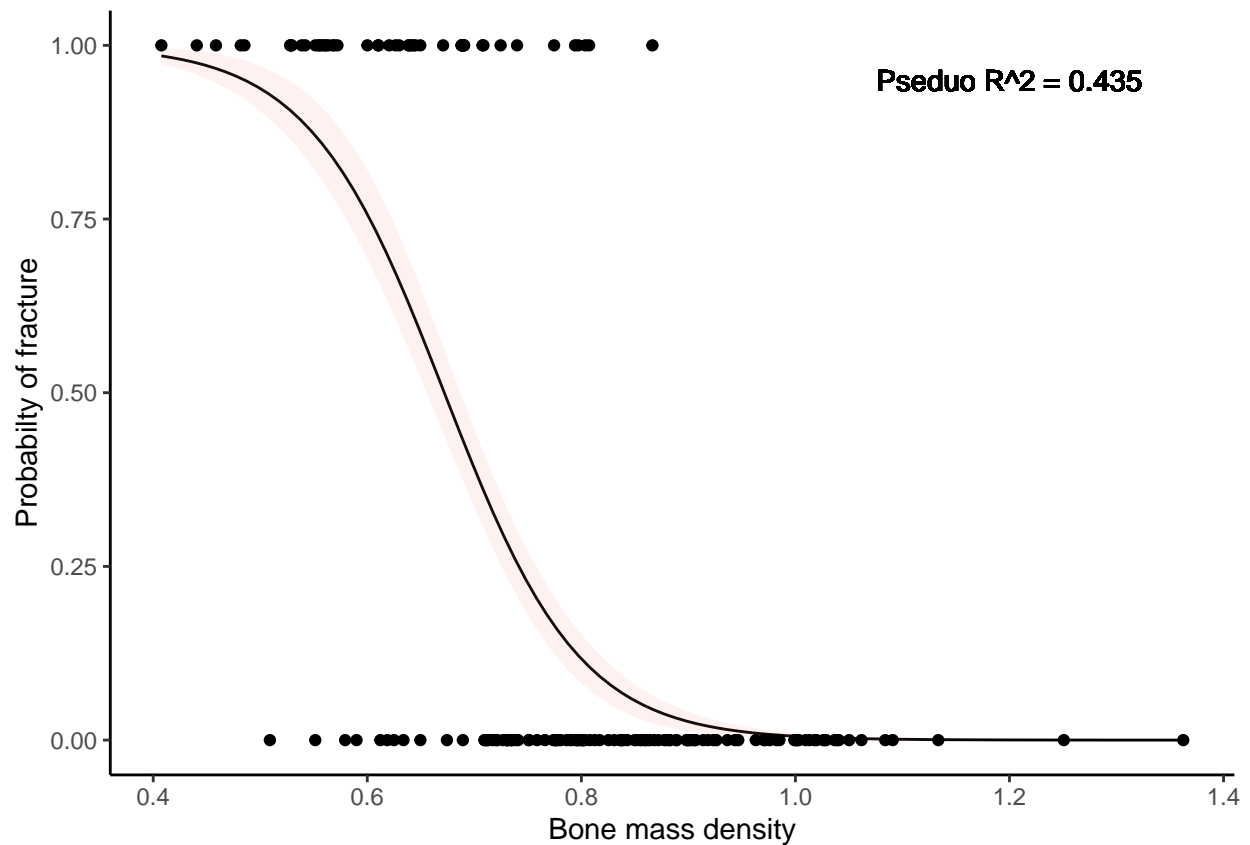
## multiple predictor

Now let ask another question; if the odds of having fracture different in male and female along with bone mass density.

Let make a exploratory graphs to check this

```r
exp_bmd <-bone_fract%>%group_by(sex)%>%summarise(prop_fract=sum(fracture)/n(),
                                                count=n())

print(exp_bmd)
```

```
## # A tibble: 2 x 3
##   sex   prop_fract count
##   <chr>      <dbl> <int>
## 1 F          0.301    83
## 2 M          0.291    86
```

```r
##also check if bmd is very different
ggplot(data = bone_fract,aes(x=bmd,fill=sex))+geom_histogram(bins = 10,
                                                alpha=0.5)
```
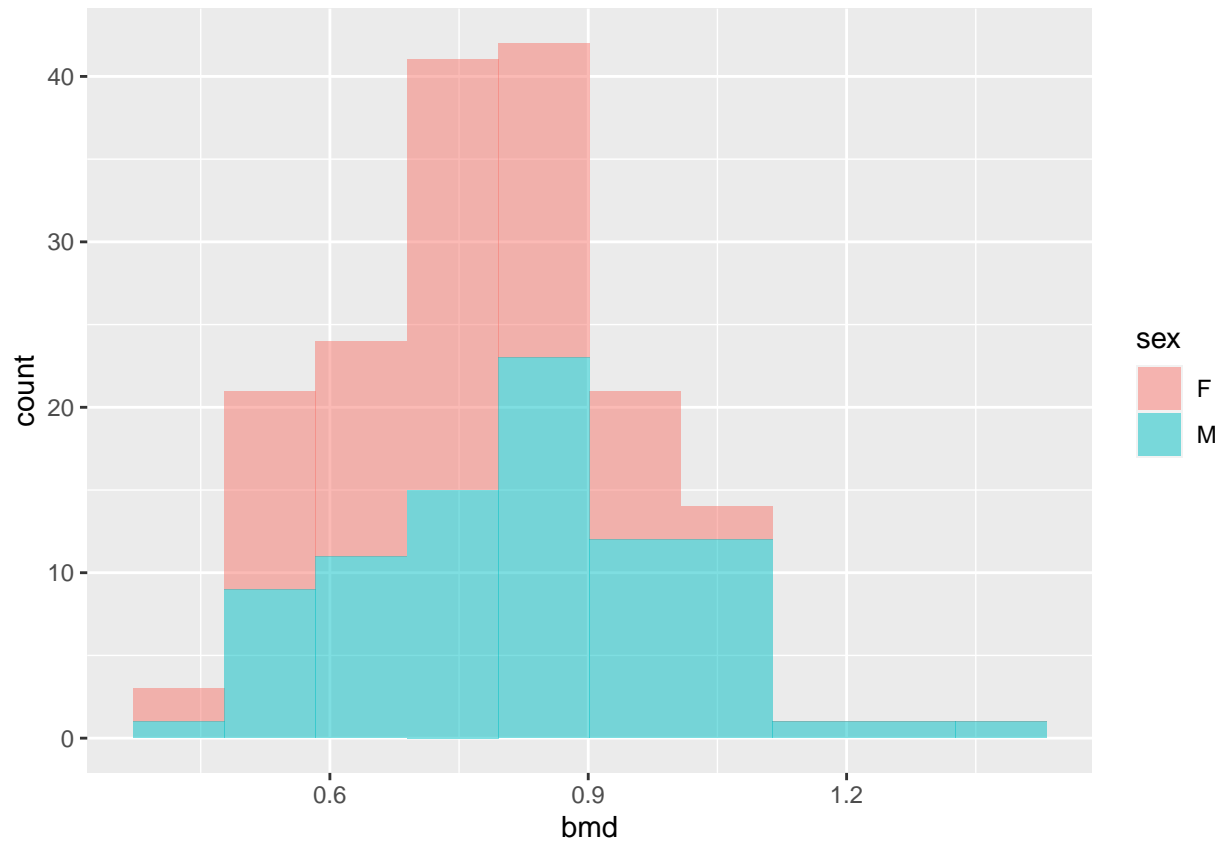
based on visual exploration we see male and female have different proportion of male and female patient have fracture. However, one examination of there bone mass density we did not see the difference. Let model this will two predictor

```
mod_gender <-  glm(fracture~bmd+sex,data = bone_fract,family = "binomial")
par(mfrow=c(2,2))
plot(mod_gender)
```

## Residuals vs Fitted



## Normal Q–Q



## Scale–Location



## Residuals vs Leverage
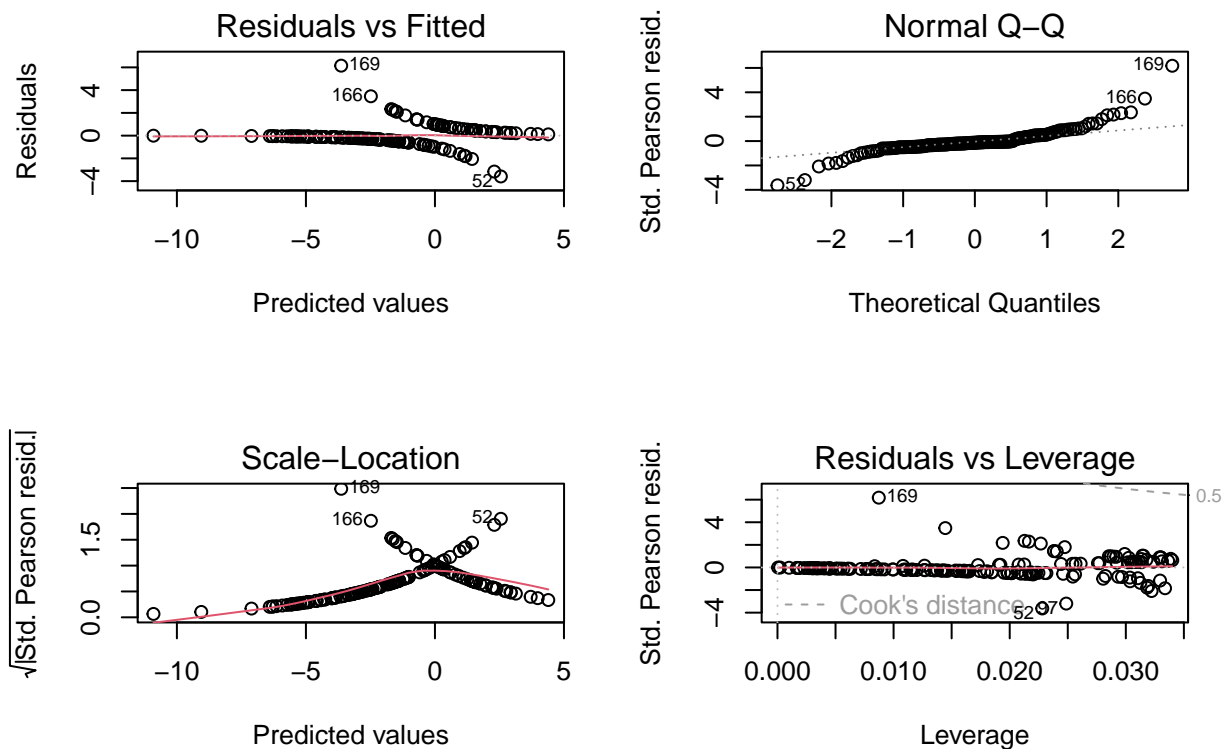


```
## Now lets see if this has improve the model or not
anova(mod_gender,mod1,test='LR')
```

```
## Analysis of Deviance Table
##
## Model 1: fracture ~ bmd + sex
## Model 2: fracture ~ bmd
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       166     112.13
## 2       167     116.02 -1  -3.8924  0.04851 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(mod_gender,mod1)
```

```
##            df      AIC
## mod_gender  3 118.1295
## mod1        2 120.0219
```

```
##very small improvement
summary(mod_gender)
```

```
##
## Call:
```

```
## glm(formula = fracture ~ bmd + sex, family = "binomial", data = bone_fract)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2934  -0.4817  -0.1901   0.3615   2.7051
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  10.7503     1.7957   5.987 2.14e-09 ***
## bmd         -16.6009     2.6045  -6.374 1.84e-10 ***
## sexM          0.9584     0.4982   1.924   0.0544 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 205.27  on 168  degrees of freedom
## Residual deviance: 112.13  on 166  degrees of freedom
## AIC: 118.13
##
## Number of Fisher Scoring iterations: 6
```

Let plot this model

```
bmd <- seq(min(bone_fract$bmd),max(bone_fract$bmd),by=0.01)
sex <- rep(unique(bone_fract$sex),length(bmd))
newdata<-data.frame(sex=sex,bmd=c(bmd,bmd))
predict_frac<-predict(mod_gender,newdata,type = "response",se.fit = T)
predict_frac<-predict_frac%>%as.data.frame()
predict_frac<-cbind(predict_frac,newdata)
##plotting
plot<-ggplot(data = predict_frac,aes(x=bmd,y=fit,col=sex))
plot+geom_line()+
  geom_ribbon(aes(ymax=fit+se.fit,ymin=fit-se.fit,fill="grey"),alpha=0.05,
              show.legend = F)+
  theme_classic()+
  labs(x="Bone mass density",y="Probabilty of fracture")+
  geom_point(data = bone_fract,aes(x=bmd,y=fracture,col=sex))
```