

Pred

Benjamin Pond

April 24, 2022

Pred

```
library(readxl)
BEI_Data_New <- read_excel("C:/Users/Benjamin Pond/Desktop/BEI_Data New.xlsx")
#View(BEI_Data_New)

df <- BEI_Data_New[1:66,]
#x1 <- df$cancer_incidence
#x2 <- df`Educational Attainment`
#x3 <- df`Pov(%)`
#ypred <- 73.76 + (0.21*x1) + (-1.31*x2) + (1.52*x3)
#View(ypred)
#summary(ypred)

#results <- cbind(df,ypred)
#results
```

```
library(readxl)
BEI_Data <- read_excel("C:/Users/Benjamin Pond/Desktop/BEI_Data_ogData.xlsx")
#View(BEI_Data_ogData)

reduced2_model <- lm(
  cancer_mortality ~
  cancer_incidence +
  educational_attainment +
  poverty,
  data = BEI_Data
)
summary(reduced2_model)
```

```
##
## Call:
## lm(formula = cancer_mortality ~ cancer_incidence + educational_attainment +
##     poverty, data = BEI_Data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -54.013 -12.989  -3.698  15.580  68.849
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   73.76399    23.36836   3.157  0.00246 **
## cancer_incidence    0.21009     0.03939   5.333 1.44e-06 ***
```

```
## educational_attainment -1.31084    0.47495 -2.760  0.00759 **
## poverty                1.52233    0.27528  5.530 6.82e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.97 on 62 degrees of freedom
## Multiple R-squared:  0.566, Adjusted R-squared:  0.545
## F-statistic: 26.95 on 3 and 62 DF,  p-value: 2.808e-11

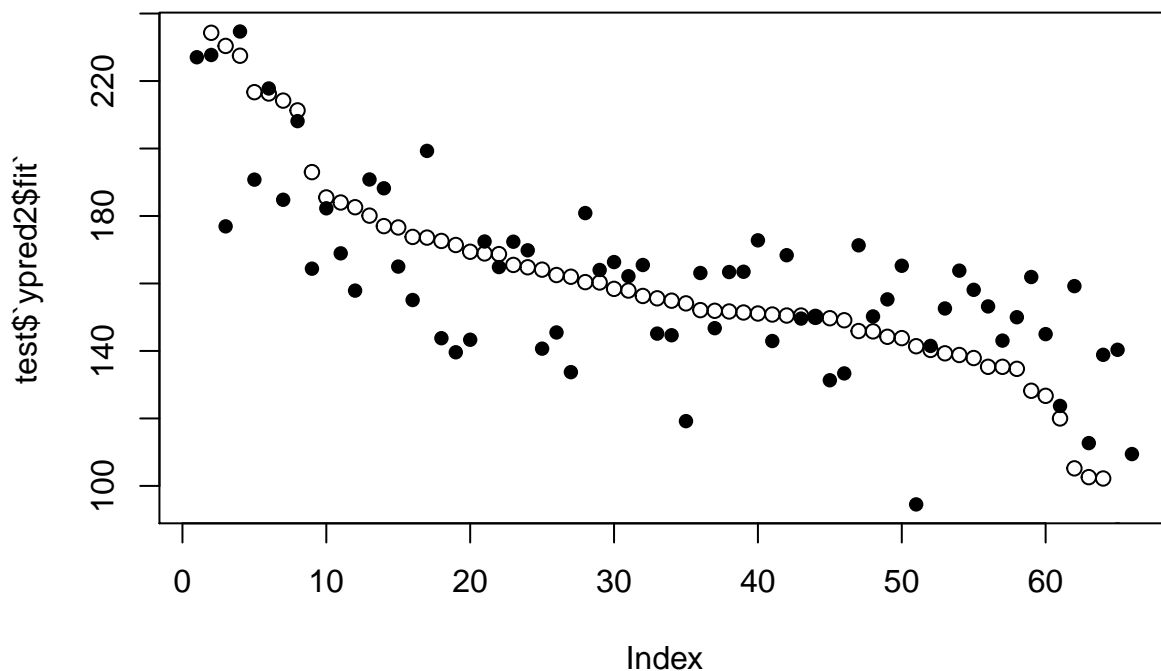
df2 <- df[,2:4]

ypred2 <- predict(reduced2_model, data = df2, se.fit = TRUE)
#ypred2

data <- cbind(df[,1], BEI_Data$cancer_mortality , ypred2$fit, ypred2$se.fit)
#data

test <- data[order(-BEI_Data$cancer_mortality),]
#test

plot(test$`ypred2$fit`,pch=16)
points(test$`BEI_Data$cancer_mortality`)
```



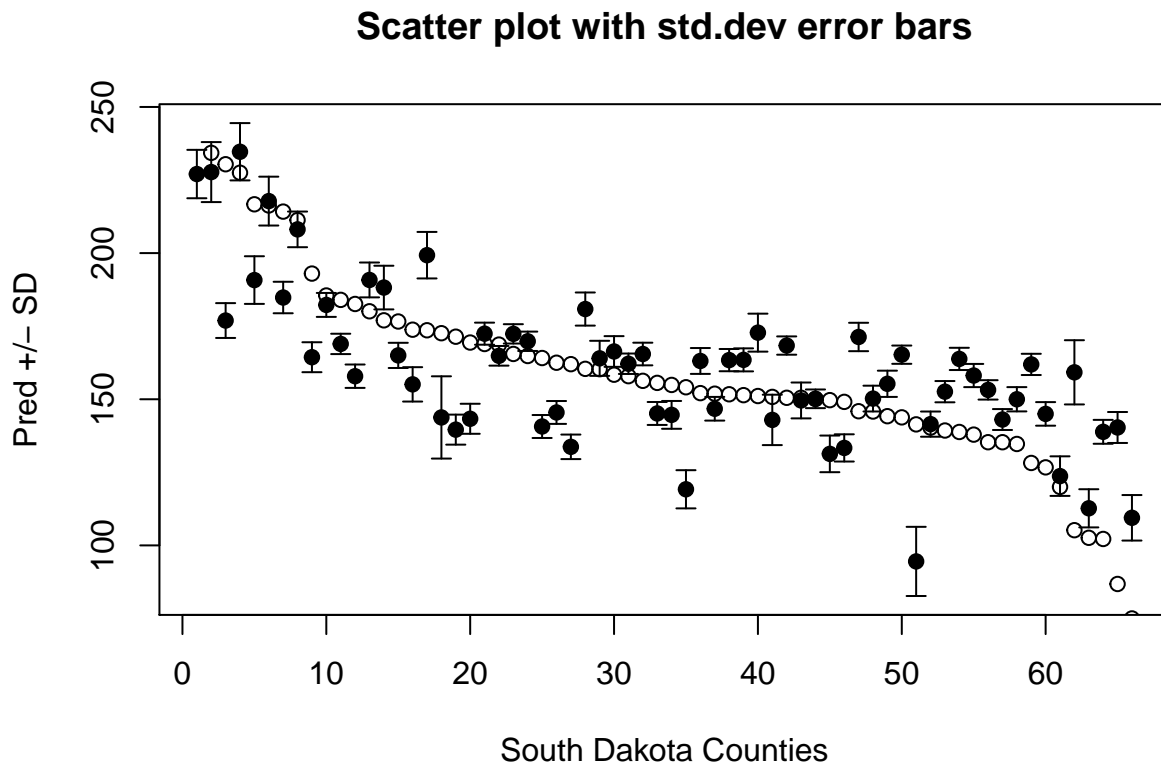
```
county <- as.factor(test$County)
mort <- test$`BEI_Data$cancer_mortality`
pred <- test$`ypred2$fit`
serr <- test$`ypred2$se.fit`
```

```

index <- 1:66

plot(index, pred,
      ylim=range(c(pred-serr, pred+serr)),
      pch=19, xlab="South Dakota Counties", ylab="Pred +/- SD",
      main="Scatter plot with std.dev error bars"
)
# hack: we draw arrows but with very special "arrowheads"
arrows(index, pred-serr, index, pred+serr, length=0.05, angle=90, code=3)
points(index, mort)

```



```

test2 <- data[order(-ypred2$fit),]

county <- as.factor(test2$County)
mort <- test2$`BEI_Data$cancer_mortality`
pred <- test2$`ypred2$fit`
serr <- test2$`ypred2$se.fit`
index <- 1:66

plot(index, pred,
      ylim=range(c(pred-serr, pred+serr)),
      pch=19, xlab="South Dakota Counties", ylab="Pred +/- SD",
      main="Scatter plot with std.dev error bars"
)
# hack: we draw arrows but with very special "arrowheads"
arrows(index, pred-serr, index, pred+serr, length=0.05, angle=90, code=3)

```

```
points(county,mort)
```

Scatter plot with std.dev error bars

