```
f=file.choose()
Taxes=read.table(f,header = T)
Taxes
    AV   SP
1  13.9 28.6
2  16.0 34.7
3  10.3 21.0
4  11.8 25.5
5  16.7 36.8
6  12.5 24.0
7  10.0 19.1
8  11.4 22.5
9  13.9 28.3
10 12.2 25.0
11 15.4 31.1
12 14.8 29.6
13 14.9 35.1
14 12.9 30.0
15 15.8 36.2
 apply(Taxes,2,mean)
  AV   SP
13.5 28.5
 apply(Taxes,2,var)
       AV        SP
 4.442857 31.611429
> attach(Taxes)
#Model 1
> Taxes1.lm=lm(SP ~ AV)
> summary(Taxes1.lm)

Call:
lm(formula = SP ~ AV)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2291 -1.0667 -0.1959  0.9770  3.0417

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -5.8121     3.0650  -1.896   0.0804 .
AV            2.5416     0.2245  11.322 4.19e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 1.771 on 13 degrees of freedom
Multiple R-squared:  0.9079,  Adjusted R-squared:  0.9008
F-statistic: 128.2 on 1 and 13 DF,  p-value: 4.187e-08

#Model 2
> Taxes2.lm=lm(SP~-1+AV)
> summary(Taxes2.lm)

Call:
lm(formula = SP ~ -1 + AV)

Residuals:
    Min      1Q  Median      3Q     Max
-2.5086 -1.6172 -0.8724  1.0767  3.5017

Coefficients:
   Estimate Std. Error t value Pr(>|t|)
AV  2.12069    0.03646   58.17   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.928 on 14 degrees of freedom
Multiple R-squared:  0.9959,  Adjusted R-squared:  0.9956
F-statistic:  3384 on 1 and 14 DF,  p-value: < 2.2e-16

#Model 3

> Taxes3.lm=lm(log(SP)~AV)
> summary(Taxes3.lm)

Call:
lm(formula = log(SP) ~ AV)

Residuals:
     Min       1Q   Median       3Q      Max
-0.06916 -0.04170 -0.01500  0.02730  0.12555

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.085146   0.107023   19.48 5.28e-11 ***
AV          0.092287   0.007839   11.77 2.63e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 0.06182 on 13 degrees of freedom
Multiple R-squared:  0.9142,  Adjusted R-squared:  0.9077
F-statistic: 138.6 on 1 and 13 DF,  p-value: 2.627e-08
```

a)

Prediction interval for Model 1

```
> predict(Taxes1.lm,data.frame(AV=13.5),interval = "predict")
   fit      lwr      upr
 28.5 24.54956 32.45044
```

Prediction interval for Model 2

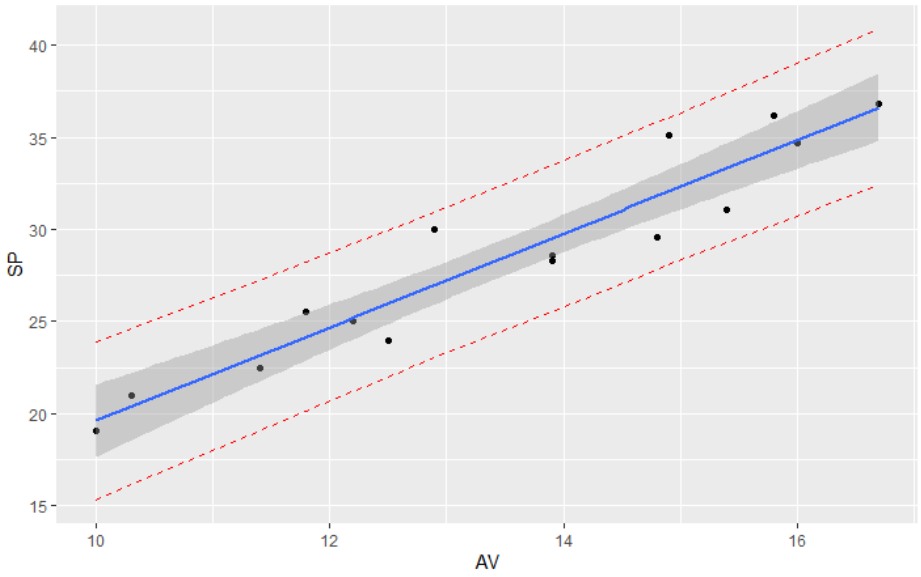
```
> predict(Taxes2.lm,data.frame(AV=13.5),interval = "predict")
      fit      lwr      upr
 28.6293 24.36216 32.89644
```

Prediction interval for Model 3

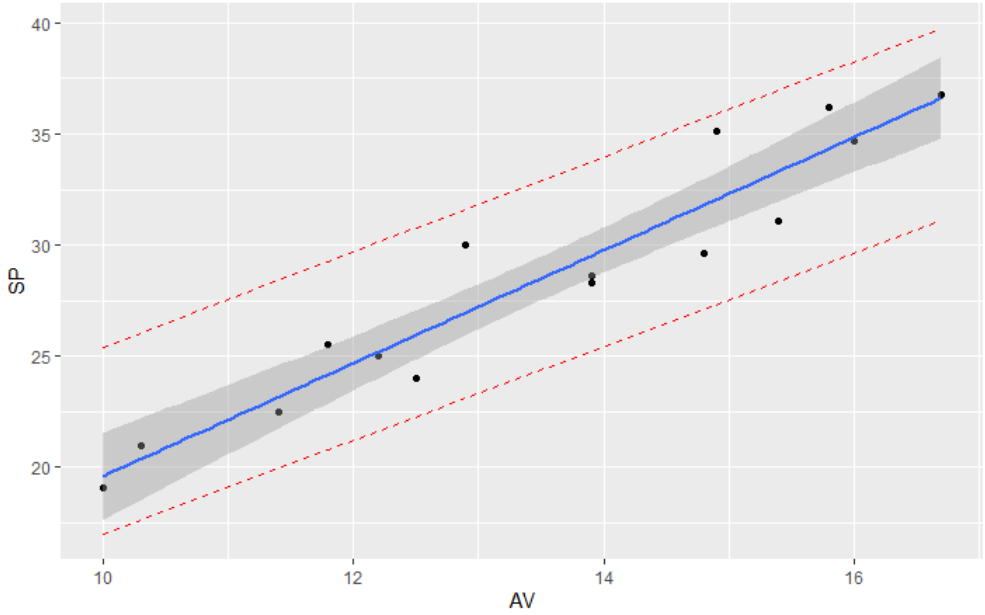
```
> predict(Taxes3.lm,data.frame(AV=13.5),interval = "predict")
       fit      lwr     upr
 3.331019 3.193077 3.46896
```

b)

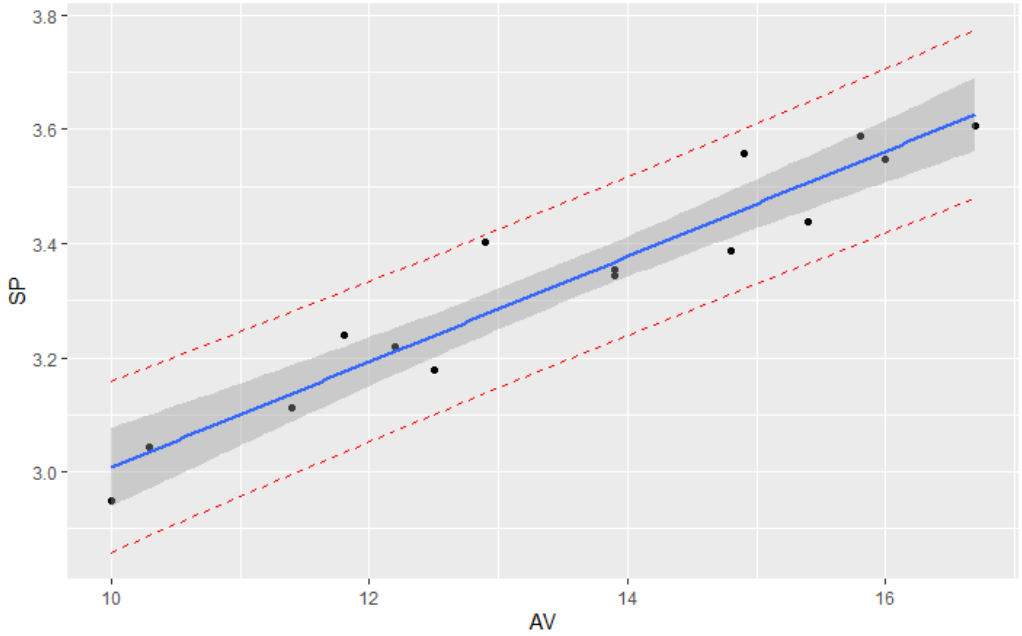
```
Taxes1.lm=lm(SP ~ AV,data = Taxes)
summary(Taxes1.lm)
predict(Taxes1.lm,data.frame(AV=13.5),interval = "predict")
Pred1=predict(Taxes1.lm,interval = "predict")
T1=cbind(Taxes,Pred1)
Plot1 <- ggplot(T1, aes(AV, SP)) +
  geom_point() +
  stat_smooth(method = lm)+
  geom_line(aes(y = lwr), color = "red", linetype = "dashed")+
  geom_line(aes(y = upr), color = "red", linetype = "dashed")
Plot1
```




```
Taxes2.lm=lm(SP~-1+AV)
summary(Taxes2.lm)
predict(Taxes2.lm,data.frame(AV=13.5),interval = "predict")
Pred2=predict(Taxes2.lm,interval = "predict")
T2=cbind(Taxes,Pred2)
Plot2 <- ggplot(T2, aes(AV, SP)) +
  geom_point() +
  stat_smooth(method = lm)+
  geom_line(aes(y = lwr), color = "red", linetype = "dashed")+
  geom_line(aes(y = upr), color = "red", linetype = "dashed")
Plot2
```



```
Taxes3.lm=lm(log(SP)~AV)
summary(Taxes3.lm)
predict(Taxes3.lm,data.frame(AV=13.5),interval = "predict")
Pred3=predict(Taxes3.lm,interval = "predict")
Taxes3=Taxes
Taxes3$SP=log(Taxes3[,2])
T3=cbind(Taxes3,Pred3)
Plot3 <- ggplot(T3, aes(AV, SP)) +
  geom_point() +
  stat_smooth(method = lm)+
  geom_line(aes(y = lwr), color = "red", linetype = "dashed")+
  geom_line(aes(y = upr), color = "red", linetype = "dashed")
Plot3
```



Bases off these graphs, I would go with the third model.The prediction intervals is more narrow and more points fall in the confidence interval compared to the other two models.

c)

Another model I would recommend would be the sqrt(X) transformation model. The sqrt(x) transformation is similar to the third model transformation and also seems to be a better fit than the first two models.

```
Taxes4.lm=lm(SP^.5~AV)
summary(Taxes4.lm)
predict(Taxes4.lm,data.frame(AV=13.5),interval = "predict")
```

```
Pred4=predict(Taxes4.lm,interval = "predict")
Taxes4=Taxes
Taxes4$SP=(Taxes4[,2])^.5
T4=cbind(Taxes4,Pred4)
Plot4 <- ggplot(T4, aes(AV, SP)) +
  geom_point() +
  stat_smooth(method = lm)+
  geom_line(aes(y = lwr), color = "red", linetype = "dashed")+
  geom_line(aes(y = upr), color = "red", linetype = "dashed")
Plot4
```