STOR 455: Homework 9 Topic 9

1. Fit a quadratic model using *Happiness* to predict GPA and construct a scatterplot of the data with the quadratic fit.

library(readr)  
Sleep = read\_csv("https://raw.githubusercontent.com/JA-McLean/STOR455/master/data/SleepStudy.csv")

## Rows: 253 Columns: 27

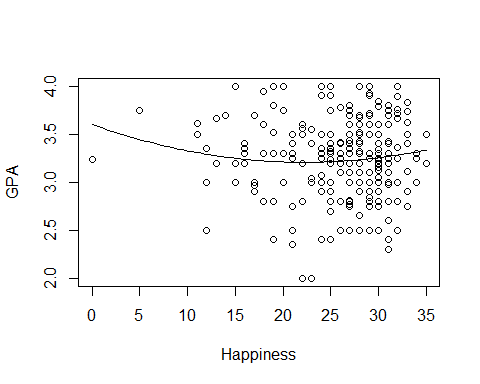
## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (5): LarkOwl, DepressionStatus, AnxietyStatus, Stress, AlcoholUse  
## dbl (22): Gender, ClassYear, NumEarlyClass, EarlyClass, GPA, ClassesMissed, ...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

mod1 = lm(GPA~Happiness+I(Happiness^2), data=Sleep)  
summary(mod1)

##   
## Call:  
## lm(formula = GPA ~ Happiness + I(Happiness^2), data = Sleep)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.20924 -0.24386 0.03966 0.27400 0.78894   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.6035793 0.2724150 13.228 <2e-16 \*\*\*  
## Happiness -0.0353191 0.0240513 -1.468 0.143   
## I(Happiness^2) 0.0007902 0.0005248 1.506 0.133   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4041 on 250 degrees of freedom  
## Multiple R-squared: 0.008996, Adjusted R-squared: 0.001068   
## F-statistic: 1.135 on 2 and 250 DF, p-value: 0.3232

plot(GPA~Happiness, data=Sleep)  
curve(summary(mod1)$coef[3,1]\*x^2 + summary(mod1)$coef[2,1]\*x + summary(mod1)$coef[1,1], add=TRUE)



1. Determine if the fit would improve if a cubic term was included.

mod2 = lm(GPA~Happiness+I(Happiness^2)+I(Happiness^3), data=Sleep)  
anova(mod1, mod2)

## Analysis of Variance Table  
##   
## Model 1: GPA ~ Happiness + I(Happiness^2)  
## Model 2: GPA ~ Happiness + I(Happiness^2) + I(Happiness^3)  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 250 40.818   
## 2 249 40.721 1 0.096534 0.5903 0.443

Doing a nested F-test, we have a p-value greater than .05 so adding the cubic term would not significantly improve the model.