PSY653, Unit 8, In class activity – Interpretation of nonlinear OLS models

Dataframe: sentence.csv

In Unit 9 we are going to replicate an analysis from a paper published by Irene Blair and colleagues at the University of CO. For now, we're just going to read in the data, format it a bit, and practice some of the nonlinear modeling techniques that we learned about in Unit 8. The data represents inmates in Florida prisons. You'll learn more about the data later. But for now, we will just consider two of the variables. The first is called years and it represents the individuals' sentence length in years. The second is primlev and it represents the severity of the primary crime that they were convicted of committing.

1. Start a new notebook called SentenceReplicationNotebook.Rmd.

2. Create a first level header called: Load libraries (i.e., # Load libraries). Insert a code chunk. Load the tidyverse, GGally, olsrr, and modelr.

3. Create a first level header called: Import data. Insert a code chunk. Import the sentence dataset.

4. Create a first level header called: Plot severity of crime and sentence length. Insert a code chunk. Make a scatterplot matrix of primlev and years. Notice that years is highly skewed to the right. And, the relationship between severity and years appears to be nonlinear.

5. Create a first level header called: Format the variables. Insert a code chunk. Use the mutate function to add the following new variables:

    a. lnyears, which equals the natural log of years
    b. primlev0, which equals primlev minus 1, so that the lowest score for primlev0 is 0.
    c. primlev02, which equals the square of primlev0

6. Create a first level header called: Plot severity of crime and log of sentence length. Insert a code chunk. Make a scatterplot matrix of primlev0 and lnyears. Please notice how things have changed.

7. Create a first level header called: Create plot of primlev0 and lnyears with loess smooth. Insert a code chunk. Create a scatterplot of primlev0 (on x axis) and lnyears (on y axis), and ask for a loess smoother. Take a look at the plot and describe what you see.

8. Create a first level header called: Create plot of primlev0 and lnyears with straight line. Insert a code chunk. Create a scatterplot of primlev0 (on x axis) and lnyears (on y axis), and ask for the best fit straight line. How well does the straight line fit the data?

9. Create a first level header called: Create plot of primlev0 and lnyears with quadratic function. Insert a code chunk. Create a scatterplot of primlev0 (on x axis) and lnyears (on y axis), and ask for the quadratic function. Take a look at the plot. Does the quadratic function seem to provide a better fit?

10. Create a first level header called: Explore models for primlev0 and lnyears. And then a second level header called: Estimate a linear model. Insert a code chunk. Call the model exp_lin and regress lnyears on primlev0. Study the results.

11. Create a second level header called: Estimate a quadratic model. Insert a code chunk. Call the model exp_quad and regress lnyears on primlev0 and primlev02. Study the results.

12. Interpret this quadratic model. Create a third level header called: Interpret quadratic model. And then a fourth level header called: Calculate the changing slope of primlev0. Insert a code chunk. Use the formula for calculating the slope drawn tangent to the parabola when primlev0 = 3, 6, and 9. Take a look at the quadratic plot that you created in Step 9, notice where these slopes would be represented.

13. These slopes represent the expected change in lnyears for a one unit increase in primlev0. If we want to know the expected change in years (rather than lnyears), then we need to use the ln.y function from Unit 8. Create a fourth level header called: Interpret regression slopes in terms of percent change in years. Insert a code chunk. Copy in the ln.y

function, and then calculate the slope to represent percent change in years for a one unit increase in x when primlev0 = 3, 6 and 9.

14. Create a fourth level header called: Calculate the predicted length of sentence in lnyears and in years. Insert a code chunk. Use the predgrid function to predict sentence length in lnyears and in years when primlev0 equals 3, 6, and 9.

15. Create a fourth level header called: Plot the model fitted relationship. Insert a code chunk. So, here's a real challenge. Go back through the Unit 8 examples and figure out how to plot the relationship between sentence length in years and severity of the crime that is represented by our fitted model (called exp_quad).

````
---
title: 'R Notebook to Replicate Blair et al. 2004 Psych Science '
output:
  html_document:
    df_print: paged
---

# Load libraries
```{r}

library(tidyverse)
library(olsrr)
library(GGally)
library(modelr)

```

# Import data
```{r}

sentence <- read_csv("sentence.csv")

```

# Plot severity of crime and sentence length
```{r}

scatterplot <- ggpairs(sentence, columns = c("primlev", "years"),
        upper = list(continuous = wrap("cor", size=3)),
        title = "Bivariate relationship of sentence length and severity of crime")
print(scatterplot, progress=FALSE )

```

# Format the variables
```{r}

sentence <- mutate(sentence,
                   lnyears = log(years),
                   primlev0 = primlev - 1,
                   primlev02 = primlev0^2)

```

# Plot severity of crime and log of sentence length
```{r}

scatterplot <- ggpairs(sentence, columns = c("primlev", "lnyears"),
        upper = list(continuous = wrap("cor", size=3)),
        title = "Bivariate relationship of lnyears and severity of crime")
print(scatterplot, progress=FALSE )

```
````

Create plot of primlev0 and lnyears with loess smooth
```{r}

ggplot(data = sentence, aes(x = primlev0, y = lnyears)) +
  geom_point() +
  geom_smooth(method = "loess", se = FALSE) +
  labs(title = "Bivariate plot with loess smooth")

```

Create plot of primlev0 and lnyears with straight line
```{r}

ggplot(data = sentence, aes(x = primlev0, y = lnyears)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Bivariate plot with straight line")

```

Create plot of primlev0 and lnyears with quadratic function
```{r}

ggplot(data = sentence, aes(x = primlev0, y = lnyears)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, formula = y ~ poly(x, 2)) +
  labs(title = "Bivariate plot with quadratic function")

```

Explore models for primlev0 and lnyears
Estimate a linear model
```{r}

exp_lin <- lm(data=sentence, lnyears ~ primlev0)
ols_regress(exp_lin)

```

Estimate a quadratic model
```{r}

exp_quad <- lm(data=sentence, lnyears ~ primlev0 + primlev02)
ols_regress(exp_quad)

```

Interpret quadratic model

Calculate the changing slope of primlev0

```{r}

slp_primlev0_3 <- exp_quad$coefficients["primlev0"] + (2*exp_quad$coefficients["primlev02"]*3)
slp_primlev0_6 <- exp_quad$coefficients["primlev0"] + (2*exp_quad$coefficients["primlev02"]*6)
slp_primlev0_9 <- exp_quad$coefficients["primlev0"] + (2*exp_quad$coefficients["primlev02"]*9)

slp_primlev0_3
slp_primlev0_6
slp_primlev0_9

```

Interpret regression slopes in terms of percent change in years

```{r}

# function to interpret regression slope for ln transformed y, original x
ln.y <- function(slope, x_chg) {
  new_slope <- 100 * (exp(slope * x_chg)-1)
  return(new_slope)
}

ln.y(slope = slp_primlev0_3, x_chg = 1)
ln.y(slope = slp_primlev0_6, x_chg = 1)
ln.y(slope = slp_primlev0_9, x_chg = 1)

```

Calculate the predicted length of sentence in lnyears and in years

```{r}

predgrid <- tibble(primlev0 = c(3, 6, 9), primlev02 = primlev0*primlev0) %>%
  add_predictions(exp_quad) %>%
  mutate(pred_years = exp(pred))

predgrid

```

Plot the model fitted relationship

```{r}

# create a scatterplot of primlev0 and years
ggplot(sentence, aes(x = primlev0, y = years)) +
  geom_point() +
  scale_y_continuous(trans = "log") +
  geom_smooth(method = lm, se = FALSE, formula = y ~ poly(x, 2)) +
  coord_trans(y = "exp") +
  labs(title = "The curvilinear relationship between crime severity and sentence length",
       y = "Sentence length in years", x = "Severity of primary crime")

```