New Predictors From Old

READING:	3.4
EXERCISES:	NONE
ASSIGNED:	HW 9
PRODUCER:	DR. MARIO

IMG CREDIT: ALEX RIEGERT-WATERS

Motivation

- The Set of Possible Predictors in a Model May Be **Larger Than** the Number of Variables in Our Data
- We May Want to Consider **Interactions** (Example: Plant Growth)
- We May Want to Consider Polynomial Terms to Capture **Nonlinearity**
- We May Want to Create Metrics Off Our Variables
    - Ratios, Differences, Lags, etc.

Regression Model With Interaction

- Model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon$$

- Interaction Variable is a **Product** of Two Other Variables

- Interaction Variable Allows the Slope of Each of the Individual Variables to be Dependent on the Value of the Other Variable

$$Y = \beta_0 + \boxed{(\beta_1 + \beta_3 X_2)} X_1 + \beta_2 X_2 + \epsilon$$

$$Y = \beta_0 + \beta_1 X_1 + \boxed{(\beta_2 + \beta_3 X_1)} X_2 + \epsilon$$

Example: Predicting Weight of Perch

- Data **Perch** in **Stat2Data** Package
- Response: *Weight*
- Predictors: *Length & Width*
- Interaction: *Length X Width*

```
library(Stat2Data)
data("Perch")
head(Perch)
```

```
##   Obs Weight Length Width
## 1 104    5.9    8.8   1.4
## 2 105   32.0   14.7   2.0
## 3 106   40.0   16.0   2.4
## 4 107   51.5   17.2   2.6
## 5 108   70.0   18.5   2.9
## 6 109  100.0   19.2   3.3
```

Example: Predicting Weight of Perch

- Model:

$$Weight = \beta_0 + \beta_1 Length + \beta_2 Width + \beta_3 (Length \times Width) + \epsilon$$

- Code:

```
mod.interact = lm(Weight ~ Length * Width, data=Perch)
summary(mod.interact)
```

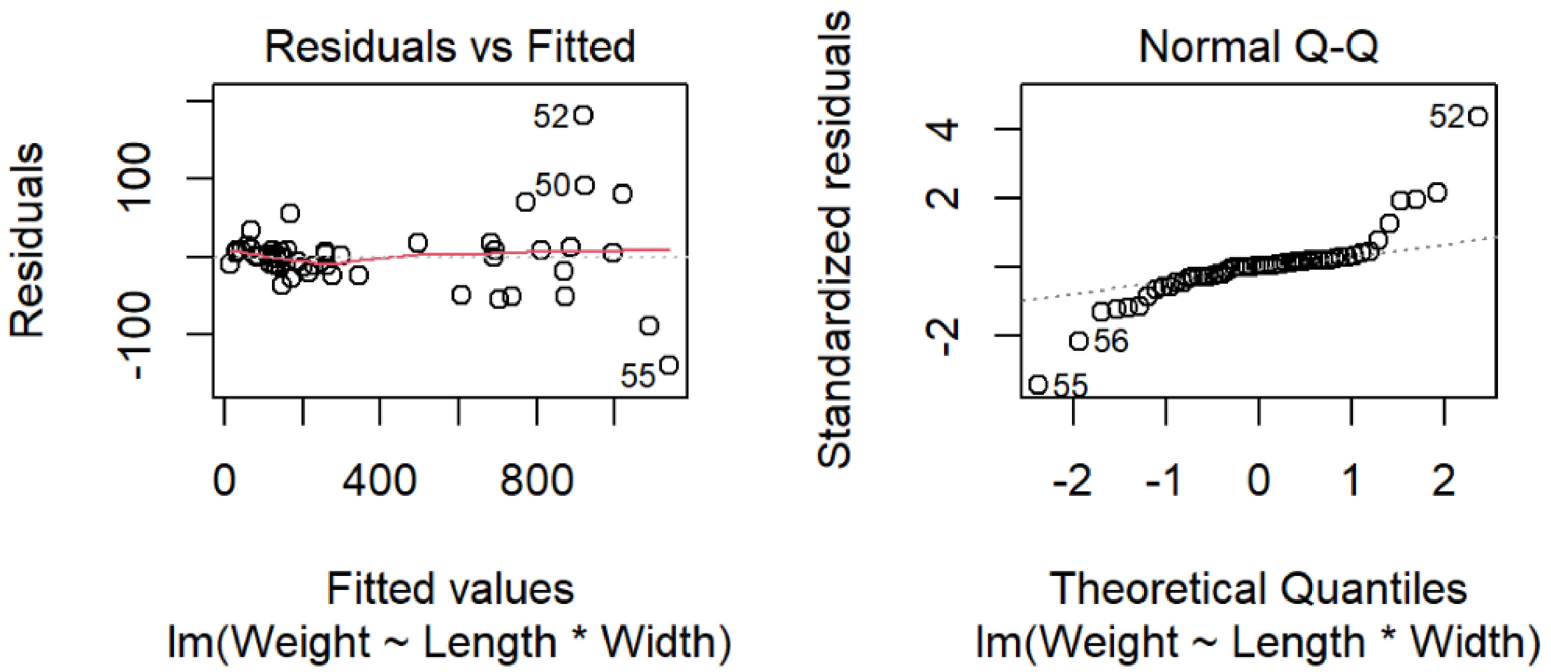
- Output:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   113.9349    58.7844   1.938    0.058 .
Length         -3.4827     3.1521  -1.105    0.274
Width         -94.6309    22.2954  -4.244 9.06e-05 ***
Length:Width    5.2412     0.4131  12.687  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 44.24 on 52 degrees of freedom
Multiple R-squared:  0.9847, Adjusted R-squared:  0.9838
F-statistic:  1115 on 3 and 52 DF,  p-value: < 2.2e-16
```

Example: Predicting Weight of Perch

- Residuals:

```
plot(mod.interact,which=c(1,2))
```



Polynomial Regression Model

- Model: $$Y = \beta_0 + \beta_1 X^1 + \cdots + \beta_k X^k + \epsilon$$
- We Need to Choose the **Degree of the Polynomial** *k*
- Polynomial Regression Models are More Difficult to **Explain**
- This Model Can Be Used to Estimate Nonlinear Relationships but It is Still a Linear Regression Model

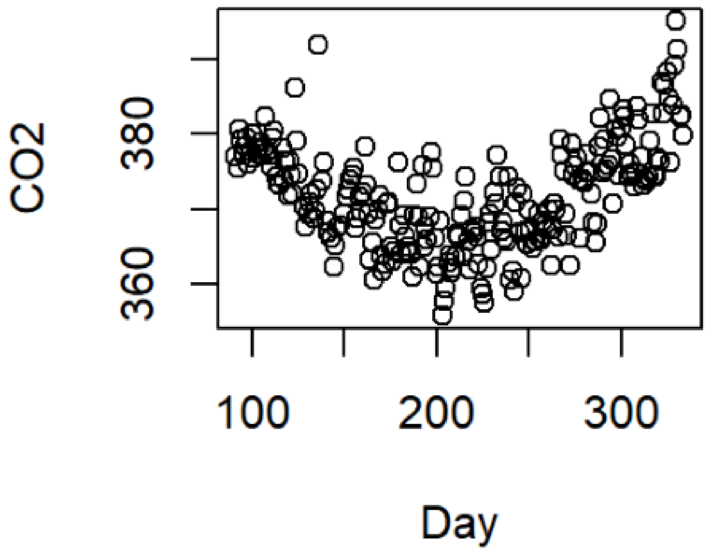
Example: Predicting CO2

- Data **CO2Germany** in **Stat2Data** Package
- Predict *CO2* Based Off *Day*
- Clear Relationship Seen

```
data("CO2Germany")
head(CO2Germany)
```

```
##       CO2  Day
## 1  377.04   91
## 2  375.52   92
## 3  380.69   93
## 4  379.21   94
## 5  377.12   95
## 6  378.38   96
```

```
library(mosaic)
plot(CO2~Day, data = CO2Germany)
```



Example: Predicting CO2

- Model:

$$CO2 = \beta_0 + \beta_1 Day + \beta_2 Day^2 + \epsilon$$

- Code:

```
poly2 = lm(CO2~Day + I(Day^2),data=CO2Germany)
summary(poly2)
```

- Output:

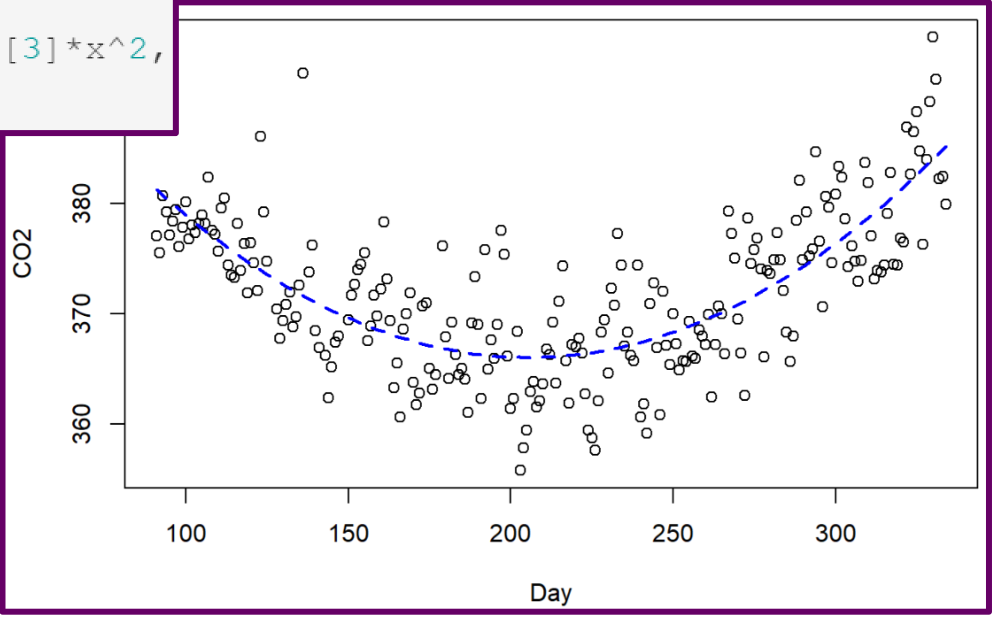
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.150e+02  2.856e+00  145.28   <2e-16 ***
Day         -4.760e-01  2.874e-02  -16.57   <2e-16 ***
I(Day^2)     1.158e-03  6.684e-05   17.32   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 4.619 on 234 degrees of freedom
Multiple R-squared:  0.5734, Adjusted R-squared:  0.5698
F-statistic: 157.3 on 2 and 234 DF,  p-value: < 2.2e-16
```

Example: Predicting CO2

- Fitted Model: $\widehat{CO2} = 414.97 - 0.476Day + 0.00116Day^2 + \epsilon$

- Visualization of Model:

```
plot(CO2~Day, data = CO2Germany)
curve(coef(poly2)[1]+coef(poly2)[2]*x+coef(poly2)[3]*x^2,
      add=TRUE,col="blue",lwd=2,lty=2)
```



Thank You

Make Reasonable Decisions