

STAT 213

Polynomials and Interactions

Colin Reimer Dawson

Oberlin College

March 7, 2018

Outline

Polynomial Regression

Interactions

Outline

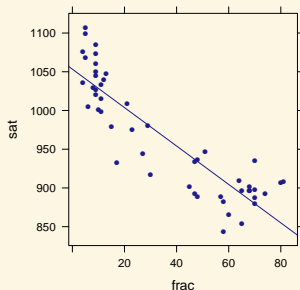
Polynomial Regression

Interactions

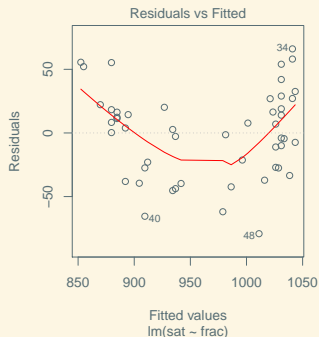
Example: State SAT Scores

```
library("mosaicData"); data("SAT")      ## sat = mean SAT score per state
slr.model <- lm(sat ~ frac, data = SAT) ## frac = % taking SAT
```

```
plotModel(slr.model)
```



```
plot(slr.model, which = 1)
```



Polynomial Regression

We can create “new” predictors from old, e.g.:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_p X^p$$

$$p = \begin{cases} 1, & \text{linear} \\ 2, & \text{quadratic} \\ 3, & \text{cubic} \\ \text{etc.} \end{cases}$$

R: Three Equivalent Methods

Method 1: Explicit Variable Creation

```
SAT.augmented <- mutate(SAT, frac.squared = frac^2)
quadratic.model <- lm(sat ~ frac + frac.squared, data = SAT.augmented)
```

Method 2: Inline transformation (note use of I())

```
quadratic.model <- lm(sat ~ frac + I(frac^2), data = SAT.augmented)
```

Method 3: Using poly() to generate polynomials

```
quadratic.model <- lm(sat ~ poly(frac, degree = 2, raw = TRUE),
                      data = SAT.augmented)
```

Call:

```
lm(formula = sat ~ frac + I(frac^2), data = SAT.augmented)
```

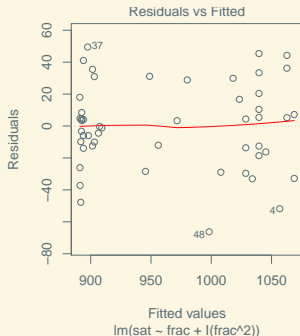
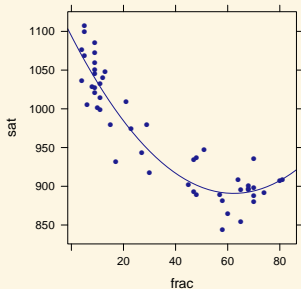
Coefficients:

(Intercept)	frac	I(frac^2)
1094.09787	-6.52850	0.05242

Example: State SAT Scores

```
## This doesn't work w/ "Method 1"  
plotModel(quadratic.model)
```

```
plot(quadratic.model, which = 1)
```



ASSESS: Do we need the quadratic term?

```
summary(quadratic.model) %>% coefficients() %>% round(digits = 2)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1094.10	9.64	113.45	0
frac	-6.53	0.73	-8.94	0
I(frac^2)	0.05	0.01	5.65	0

Selecting Polynomial Order

- Start with a higher-order model, then remove highest order term if not significant.
- Repeat until highest order term is significant.
- To be safe(ish): nested F -test between final model and highest-order model.
- Don't remove lower order terms even if nonsignificant!

Outline

Polynomial Regression

Interactions

Interaction Terms and Second-Order Models

Consider the model:

$$\text{sat} = \beta_0 + \beta_1 \cdot \text{frac} + \beta_2 \cdot \text{expend} + \beta_3 \cdot \text{frac} \cdot \text{expend} + \varepsilon$$

where `expend` is state education expenditure per pupil.

How can we interpret β_3 ? Represents change in slope relating `sat` to `expend` for *each unit increase* in `frac` (or vice versa)

Interaction Visualization

Demo

The Economic Value of a College Degree

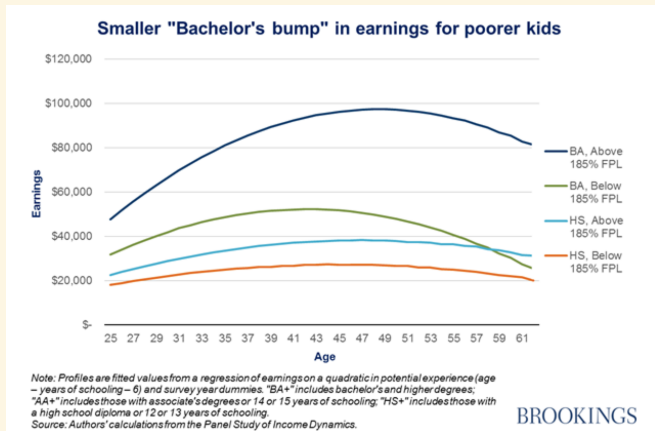


Figure: Source: <http://www.pbs.org/newshour/making-sense/if-you-grew-up-poor-your-college-degree-may-be-worth-less/>