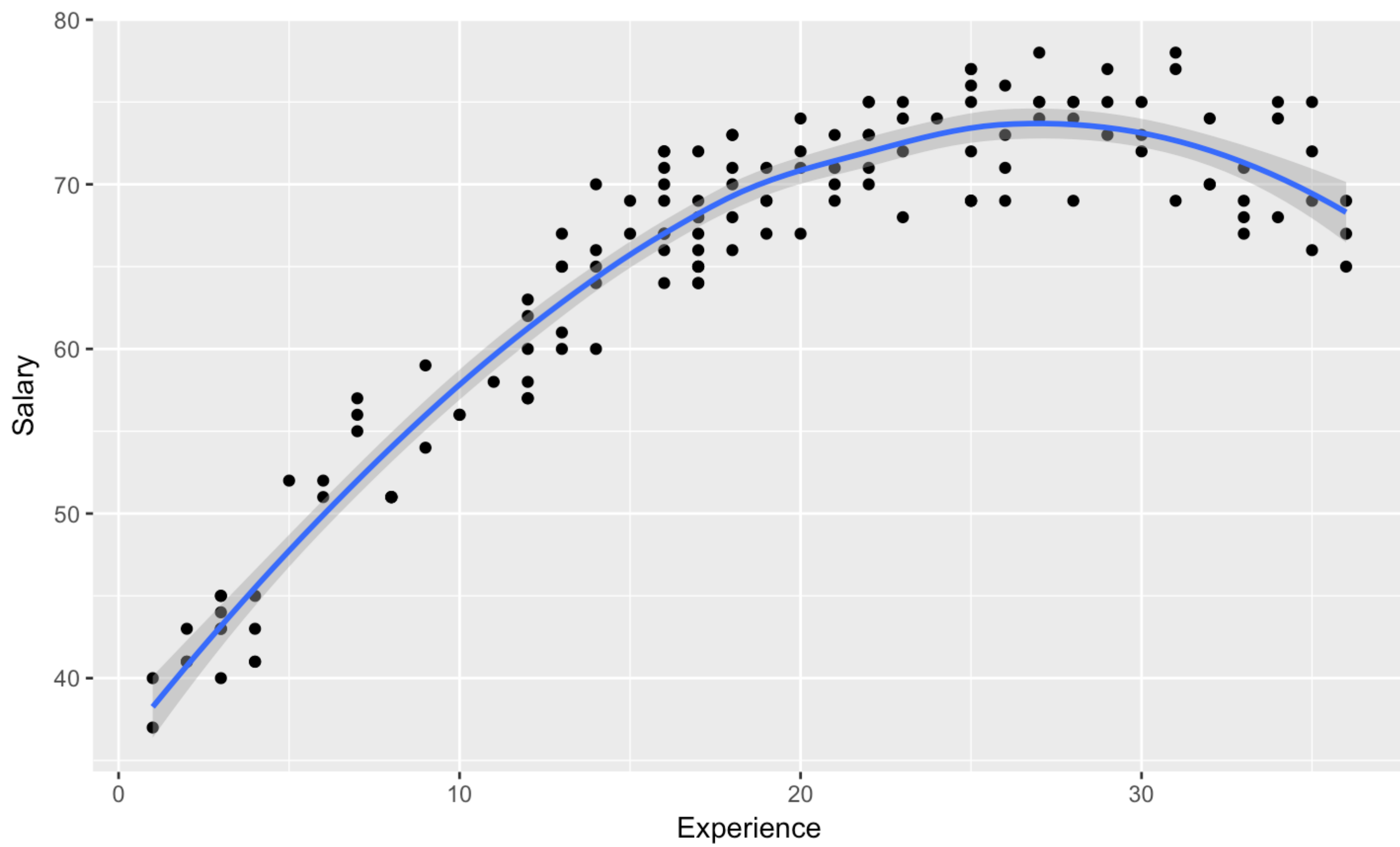
Polynomial Regression

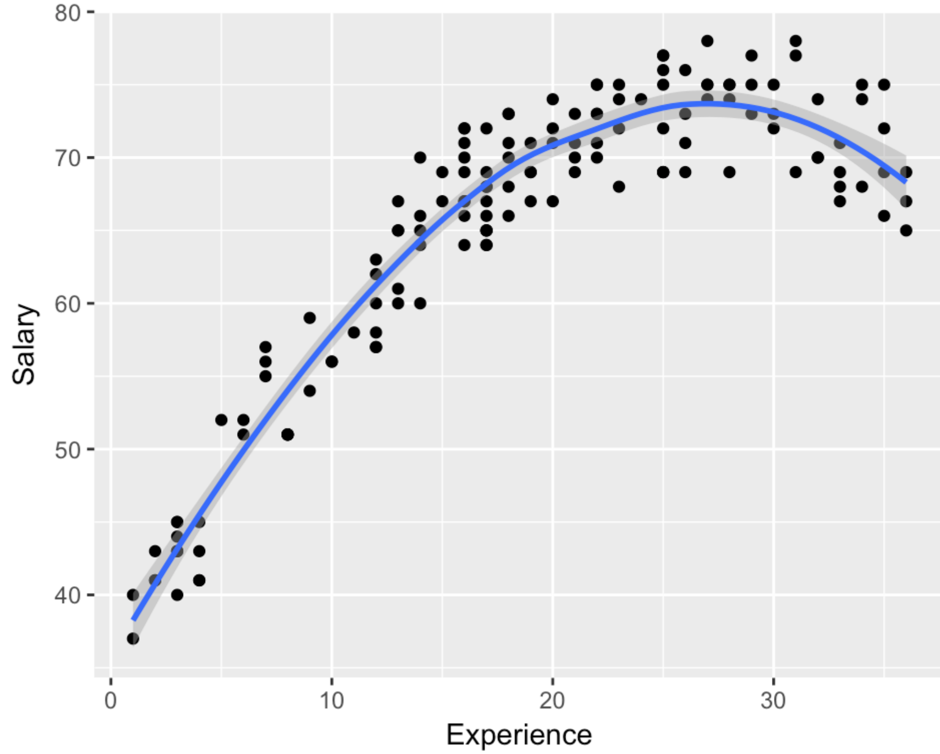
Math 430, Winter 2017

Modeling salary

What if the relationship between salary and years of experience wasn't linear?



Polynomial regression



A quadratic model seems sensible.

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$$

Polynomial regression in R

```
quad.mod <- lm(Salary ~ Experience + I(Experience^2), data = profsalary)
summary(quad.mod)
```

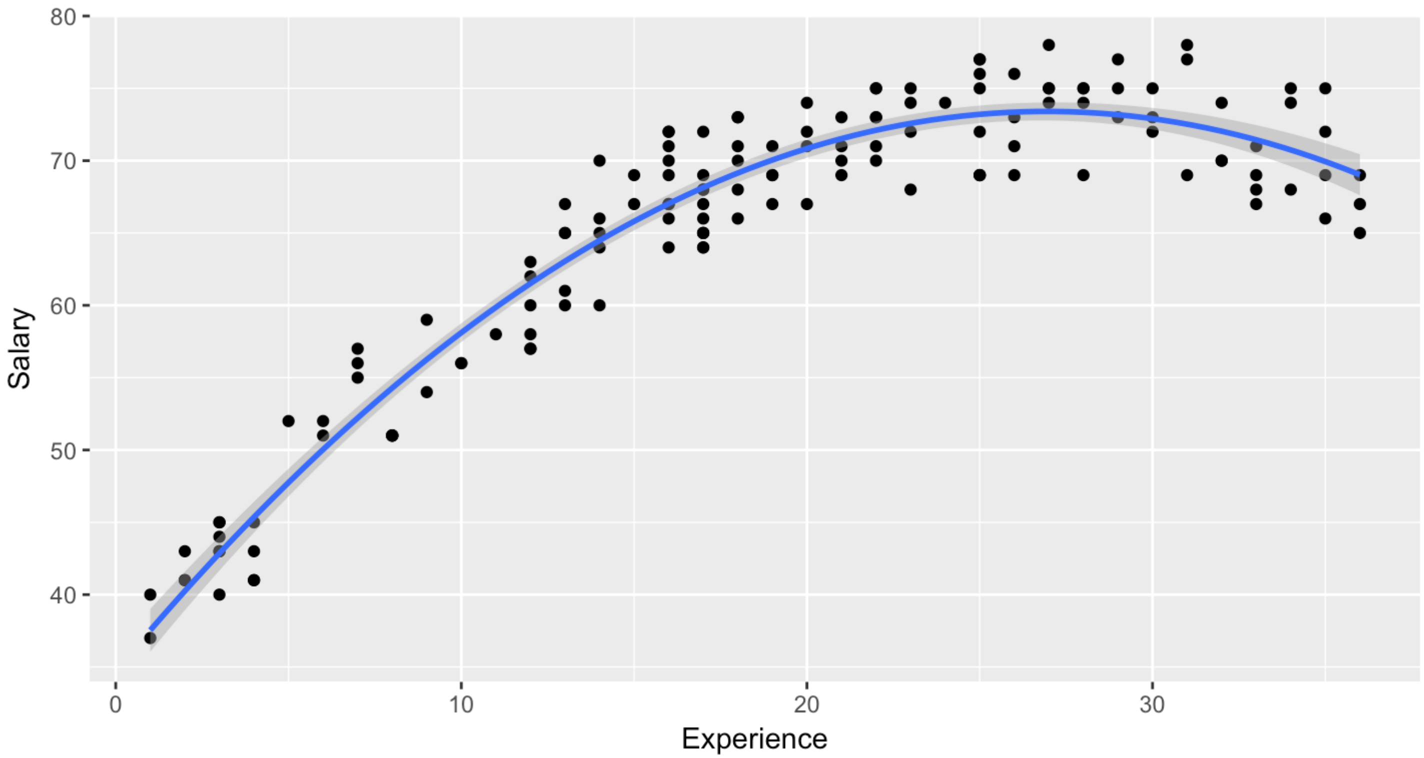
```
##
## Call:
## lm(formula = Salary ~ Experience + I(Experience^2), data = profsalary)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5786 -2.3573  0.0957  2.0171  5.5176
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     34.720498   0.828724   41.90   <2e-16 ***
## Experience       2.872275   0.095697   30.01   <2e-16 ***
## I(Experience^2) -0.053316   0.002477  -21.53   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.817 on 140 degrees of freedom
## Multiple R-squared:  0.9247, Adjusted R-squared:  0.9236
## F-statistic: 859.3 on 2 and 140 DF,  p-value: < 2.2e-16
```

Interpreting coefficients

Interpreting coefficients

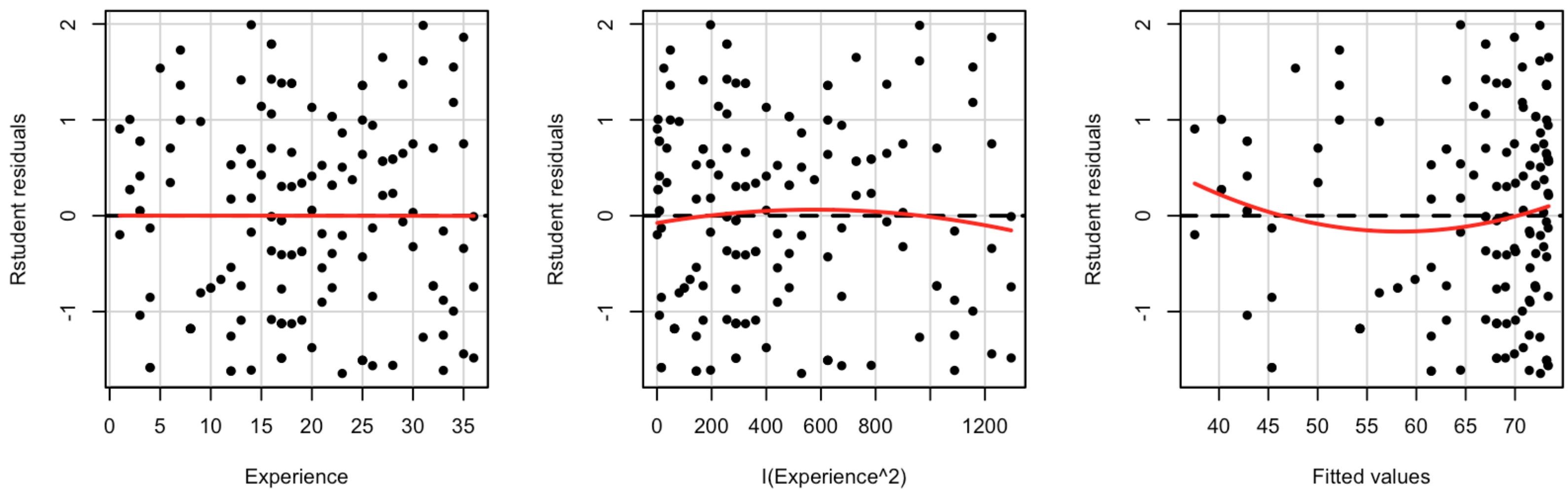
Polynomial regression

```
ggplot(data = profsalary, aes(x = Experience, y = Salary)) +
  geom_point() +
  geom_smooth(method = "lm", formula = y ~ poly(x, 2))
```



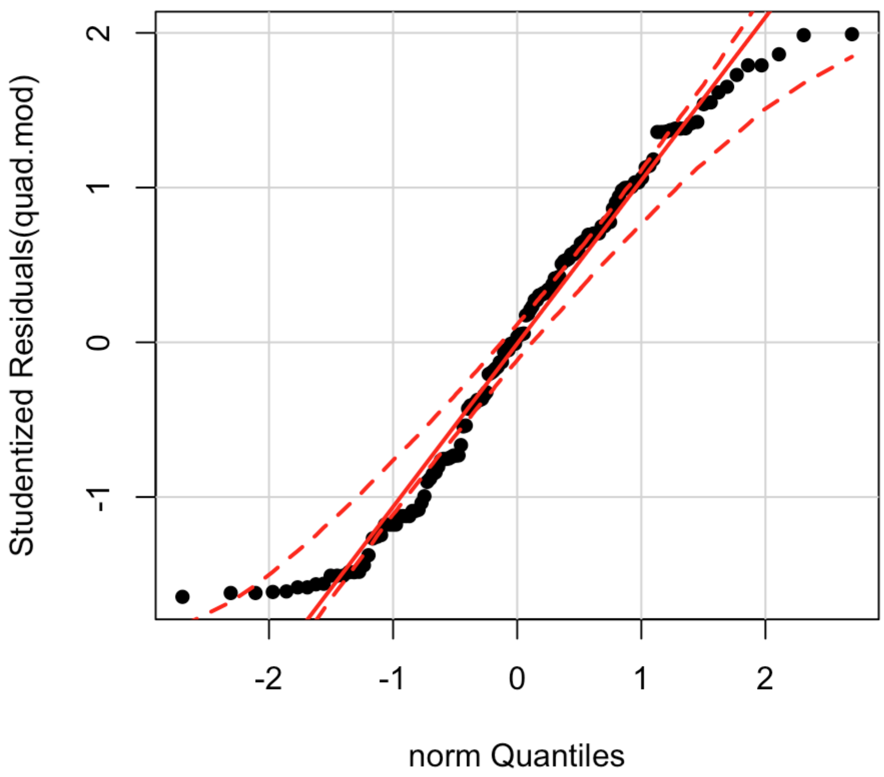
Model checking

```
residualPlots(quad.mod, type = "rstudent", pch = 16)
```



Model checking

```
qqPlot(quad.mod, pch = 16, dist = "norm", reps = 5000)
```



Beyond second-order models

Polynomials for one predictor

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \varepsilon_i$$