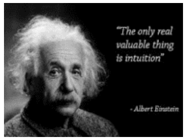
Transformation #2:
Polynomials

LI(✘) – Relationship between Y and X's are not linear



Polynomials: Intuition

- Polynomial transformations are very useful when the relationship between the X's and Y are suspected to be **non-linear**
- We cover **non-linear models** in depth later on, so we will only discuss this briefly here
- Generally, a quadratic model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ is the preferred polynomial if the data is **curvilinear** or **with 1 peak/valley**.
- A **cubic** model is preferred with **2 peaks/valleys**, etc.
- The more **wavy** the relationship the higher the polynomial
- The problem is that **high polynomials** are **difficult to interpret** and they tend to **"over-identify"** the model and do not generally perform well with new data, especially at **both ends** of the curve
- **Spline** and **piecewise** models (covered later in the class) generally perform better than high polynomials.
- Quadratic and cubic transformations are the **most popular** polynomials.

R Tips

`lm.fit2=lm(y~x+I(x^2)+I(x^2),data=dataName)` → To include a quadratic, cubic, etc. term. The `I()` (i.e., "identity) function is needed because the `^` symbol has other interpretations in R

`lm.fit5=lm(y~poly(x,5), data=dataName)` → the `poly()` function can be used to include several polynomial terms. In this example the `poly()` function add x^2 up to x^5 into the model

You can use the anova() function to compare models:

```
anova(lm.fit2,lm.fit5)
```

Kogod School of Business