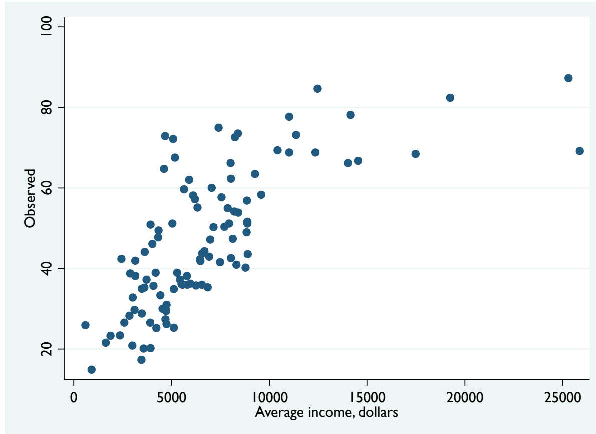


What is the relationship between income and prestige?

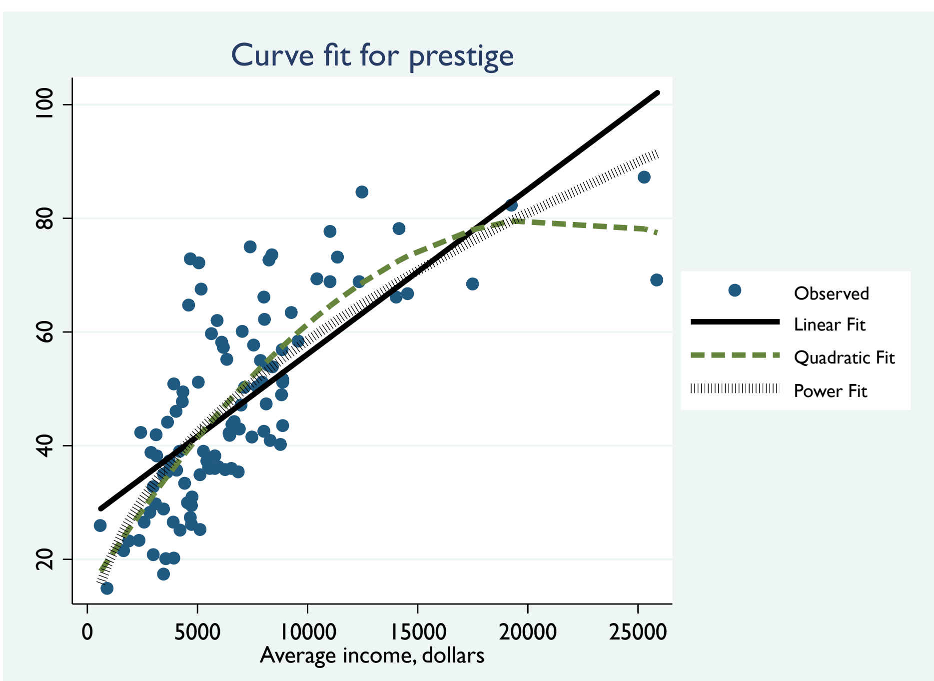
Some popular options:
Linear: Y = b0 + (b1 * X)
Logarithmic: Y = b0 + (b1 * ln(X))
Inverse: Y = b0 + (b1 / X)
Quadratic: Y = b0 + (b1 * X) + (b2 * X^2)
Cubic: Y = b0 + (b1 * X) + (b2 * X^2) + (b3 * X^3)
Power: Y = b0 * (X^b1) OR ln(Y) = ln(b0) + (b1 * ln(X))
Compound: Y = b0 * (b1^X) OR ln(Y) = ln(b0) + (ln(b1) * X)
S-curve: Y = e^(b0 + (b1/X)) OR ln(Y) = b0 + (b1/X)
Logistic: Y = b0 / (1 + b1 * e^(-b2 * X))
Growth: Y = e^(b0 + (b1 * X)) OR ln(Y) = b0 + (b1 * X)
Exponential: Y = b0 * (e^(b1 * X)) OR ln(Y) = ln(b0) + (b1 * X)
Vapor Pressure: Y = e^(b0 + b1/X + b2 * ln(X))
Reciprocal Logarithm: Y = 1 / (b0 + (b1 * ln(X)))
Modified Power: Y = b0 * b1^(X)
Shifted Power: Y = b0 * (X - b1)^b3
Geometric: Y = b0 * X^(b1 * X)
Modified Geometric: Y = b0 * X^(b1/X)
nth order Polynomial: Y = b0 + b1*X + b2*X^2 + b3*X^3 + b4*X^4 +b5*X^5 ...
Hoerl: Y = b0 * (b1^X) * (X^b2)
Modified Hoerl: Y = b0 * b1^(1/X) * (X^b2)
Reciprocal: Y = 1 / (b0 + b1 * X)
Reciprocal Quadratic: Y = 1 / (b0 + b1 * X + b2 * X^2)
Bleasdale: Y = (b0 + b1 * X)^(-1 / b2)
Harris: Y = 1 / (b0 + b1 * X^b2)
Exponential Association: Y = b0 * (1 - e^(-b1 * X))
Three-Parameter Exponential Association: Y = b0 * (b1 - e^(-b2 *X))
Saturation-Growth Rate: Y = b0 * X/(b1 + X)
Gompertz Relation: Y = b0 * e^(-e^(b1 - b2 * X))
Richards: Y = b0 / (1 + e^(b1 - b2 * X))^(1/b3)
MMF: Y = (b0 * b1+b2 * X^b3)/(b1 + X^b3)
Weibull: Y = b0 - b1*e^(-b2 * X^b3)
Sinusoidal: Y = b0+b1 * b2 * cos(b2 * X + b3)
Gaussian: Y = b0 * e^((-(b1 - X)^2)/(2 * b2^2))
Heat Capacity: Y = b0 + b1 * X + b2/X^2
Rational: Y = (b0 + b1 * X)/(1 + b2 * X + b3 * X^2)

Let's fit 3 models:

Linear: Prestige = b0 + (b1 * income)
Quadratic: Prestige = b0 + (b1 * income) + (b2 * income^2)
Power: Prestige = b0 * (income^b1)   OR   ln(prestige) = ln(b0) + (income * ln(X))

To fit these models, I use the Stata command: *curvefit prestige income, f(146)*

```
Curve Estimation between prestige and income
--------------------------------------------------------
    Variable |  Linear       Quadratic        Power
-------------+------------------------------------------
b0           |
       _cons |  27.141177     14.183194     .77539055       <--- coefficient
             |      11.97          4.03          2.74       <--- t-statistic
             |     0.0000        0.0001        0.0073       <--- p-value
-------------+------------------------------------------
b1           |
       _cons |   .0028968     .00615351     .46942659
             |      10.22          8.10         11.56
             |     0.0000        0.0000        0.0000
-------------+------------------------------------------
b2           |
       _cons |               -1.433e-07
             |                    -4.56
             |                   0.0000
-------------+------------------------------------------
Statistics   |
           N |        102           102           102
        r2_a |    .506201     .58786733     .94848615
--------------------------------------------------------
                                          legend: b/t/p
```



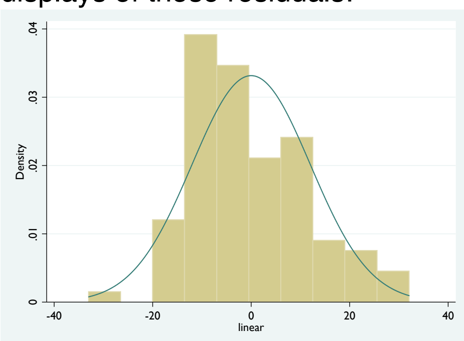
Write out these models and interpret them.

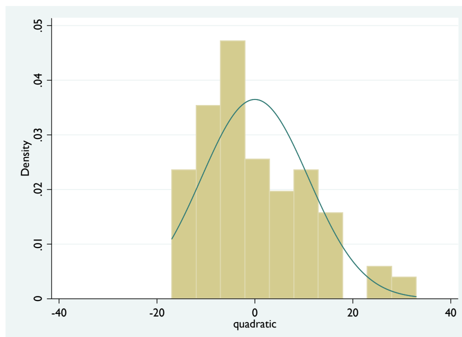
Which model is best?

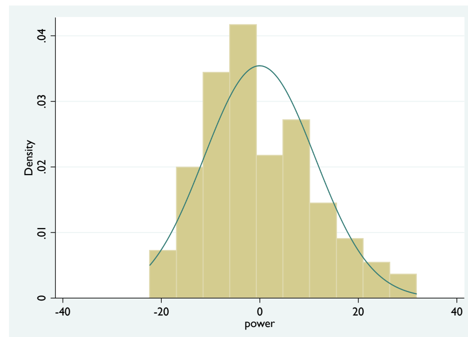
I then had Stata compute the predicted prestige values from each model.

#	Title	Income (X_1)	Prestige (Y)	Linear Prediction	Quadratic Prediction	Power Prediction
1	Physicians	25308	87.2	100.4534	78.13331	90.47463
2	University Professors	12480	84.6	63.29323	68.65997	64.92210
…	…	…	…	…	…	…
101	Janitors	3472	17.3	37.19886	33.82073	35.60925
102	Newsboys	918	14.8	29.80044	19.71135	19.07030
	Means	**6797.90**	**46.833**	**46.833**	**46.834**	**46.894**
	Std. Deviations	**4245.92**	**17.204**	**12.300**	**13.283**	**12.784**

Using these predicted values, I calculated the residuals = (observed prestige) – (predicted prestige). Here are displays of those residuals:







and summary statistics

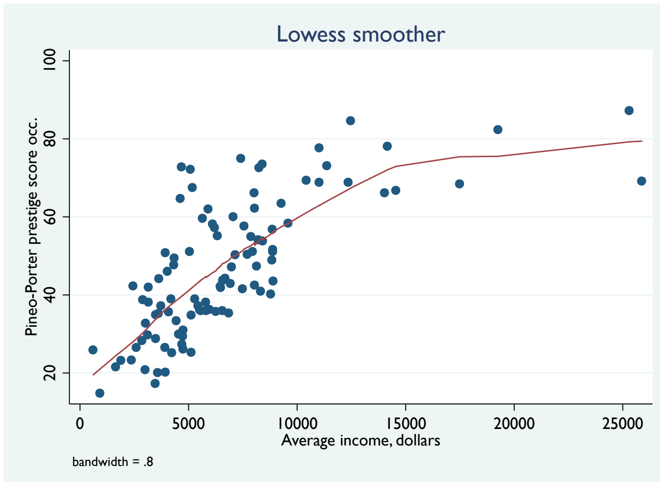
```
   stats |     linear quadratic    power
---------+------------------------------
     min | -33.00745 -16.96424 -22.32719
    mean | -1.03e-07 -.0006502 -.0607795
     p50 |  -2.37771 -2.303448 -2.641863
      sd |  12.02974  10.93496  11.26121
     max |  32.08443  32.92813  31.80861
----------------------------------------
```

Which model would you choose?

Yet another option would be to use nonparametric regression -- making no assumptions about the shape of the relationship between prestige and income. In this example, we'll use *locally weighted scatterplot smoothing*, or LOWESS regression.

LOWESS regression fits a smooth curve through our scatterplot. It does this by fitting a polynomial to subsets of the scatterplot.

Here's the LOWESS curve generated by our prestige/income data:

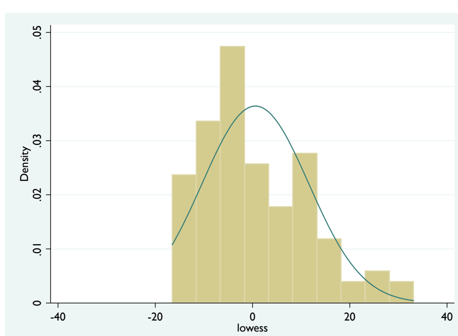


We can then have the computer use the LOWESS curve to estimate our predicted prestige values:

#	Title	Income (X_1)	Prestige (Y)	Linear Prediction	Quadratic Prediction	Power Prediction	LOWESS Prediction
1	Physicians	25308	87.2	100.4534	78.13331	90.47463	79.23667
2	University Professors	12480	84.6	63.29323	68.65997	64.92210	67.42372
…	…	…	…	…	…	…	
101	Janitors	3472	17.3	37.19886	33.82073	35.60925	33.70652
102	Newsboys	918	14.8	29.80044	19.71135	19.07030	20.99867
	Means	**6797.90**	**46.833**	**46.833**	**46.834**	**46.894**	**46.244**
	Std. Deviations	**4245.92**	**17.204**	**12.300**	**13.283**	**12.784**	**12.834**

Then we can calculate the residuals and display them:

```
   stats |    lowess
---------+----------
     min | -16.51583
    mean |  .5895998
     p50 | -1.875635
      sd |  10.96312
     max |     33.13
--------------------
```



LOWESS regression requires a large sample size and it does not generate a simple function between our independent and dependent variable.