25. Polynomial Regression

Readings: Kleinbaum, Kupper, Nizam, and Rosenberg (KKNR): Ch. 15

SAS: PROC REG, PROC LOESS

Homework: Homework 9 due by 11:59 pm on November 28
Final Project due by 11:59 on December 6

Overview

A) Re/Preview of Topics
B) Polynomial Models with One Variable
C) Estimate of $\sigma^2_{Y|X}$ and Lack of Fit
D) Hierarchical Modeling
E) Appendix: Other Remedies for Non-Linearity

A. Review (Lectures 23-24)/Current (Lecture 25)/Preview (Lecture 26)

Lecture 23-24:

- Categorical Predicators
    - Indicator variables
- Test of general linear hypothesis

Lecture 25:

- Polynomial Regression
    - Quadratic, cubic, quartic
- Other remedies for non-linearity

Lecture 26:

- Model Selection
    - Ways to select the “best” model

B. Polynomial Models with One Variable

A k^{th} order polynomial in one variable, *x*, is an expression of the following form:

$$y = c_0 + c_1 x + c_2 x^2 + \cdots + c_k x^k$$

in which the *c*'s and *k* (which must be a nonnegative whole number) are constants.

The statistical model is an expression of the following form:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_k X^k + \epsilon$$

This statistical model is a linear regression model because *Y* is a linear function of $\boldsymbol{\beta}$.

Polynomial models are useful:

- In situations where the analyst knows that curvilinear effects are present in the true response function.
- As approximating functions to unknown and possibly very complex nonlinear relationships.

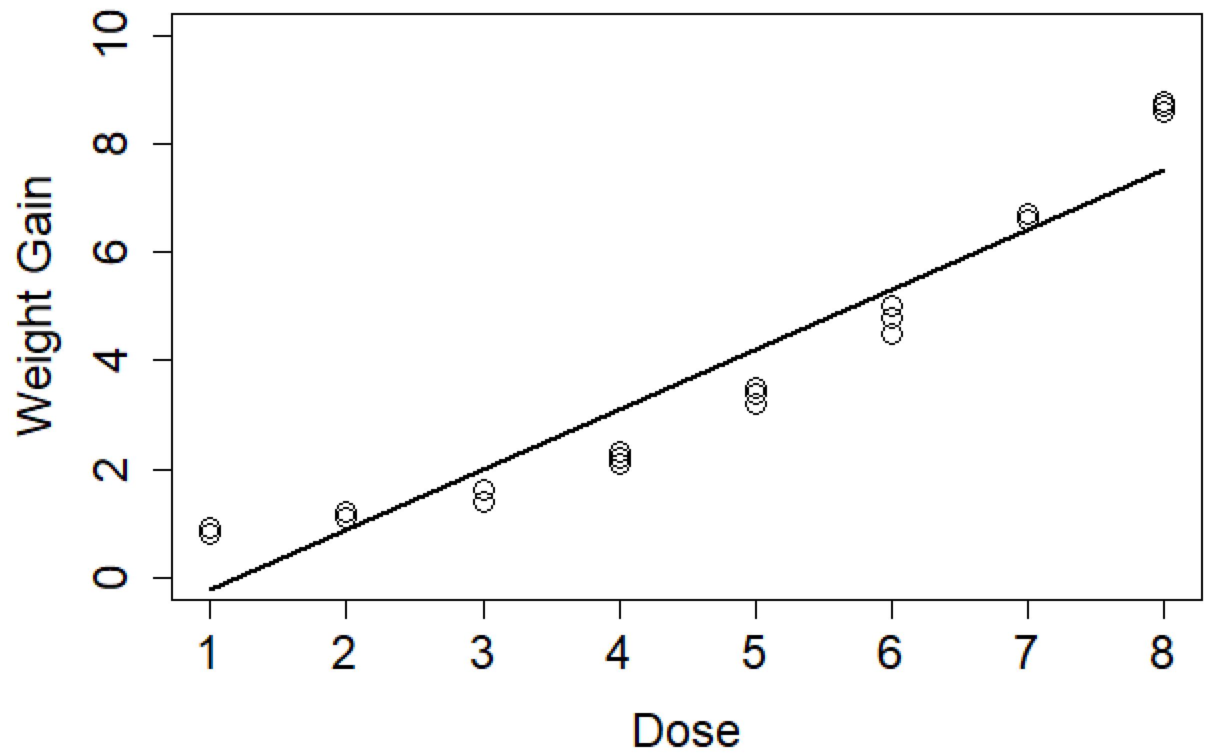
Important considerations when using polynomial models include:

- Selecting the order of the model (model selection strategy)
- Extrapolation
- Ill-conditioning

Example (Linear Model)

Example: (from KKNR) A laboratory study is undertaken to determine the relationship between the dosage (X) of a certain drug fixed by the investigator and weight gain (Y). 24 laboratory animals of the same sex, age, and size are selected and 3 animals are randomly assigned to each dose group. *(Actual data provided on slide 11, code to read in data in SAS file for Lecture 25.)*

Scatterplot:




```
PROC REG DATA=wtgain;
    MODEL wgtgain = dose;
RUN;
```

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	**155.66669**	155.66669	238.27	<.0001
Error	**22**	**14.37290**	**0.65331**		
Corrected Total	23	170.03958			

MSE contains pure error + lack of fit error

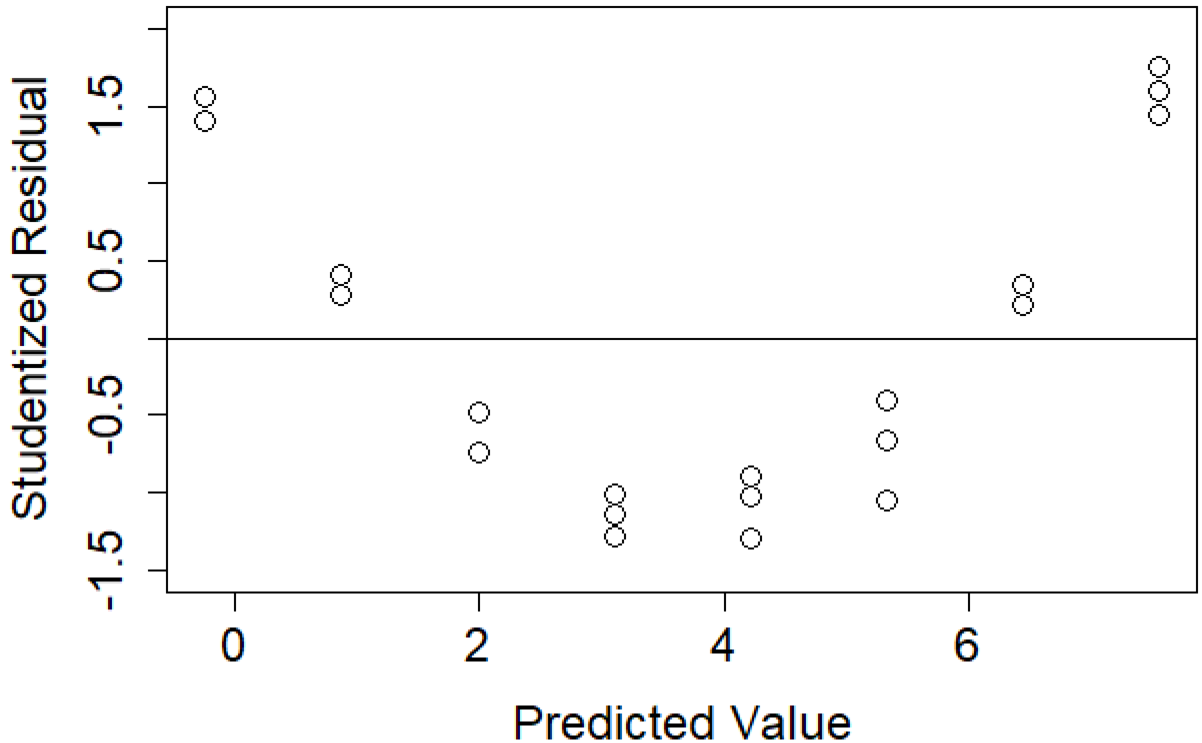
Root MSE	0.80828	**R-Square**	0.9155
Dependent Mean	3.65417	**Adj R-Sq**	0.9116
Coeff Var	22.11936		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	-1.34762	0.36362	-3.71	0.0012
dose	1	**1.11151**	**0.07201**	**15.44**	**<.0001**

Intercept: Predicted weight gain when the dose is 0. This is beyond the range of our data.

β_1: For every one unit increase in dose, weight gain increases by 1.11 units, on average.

Residual Plot:



Does a straight-line model fit the data?

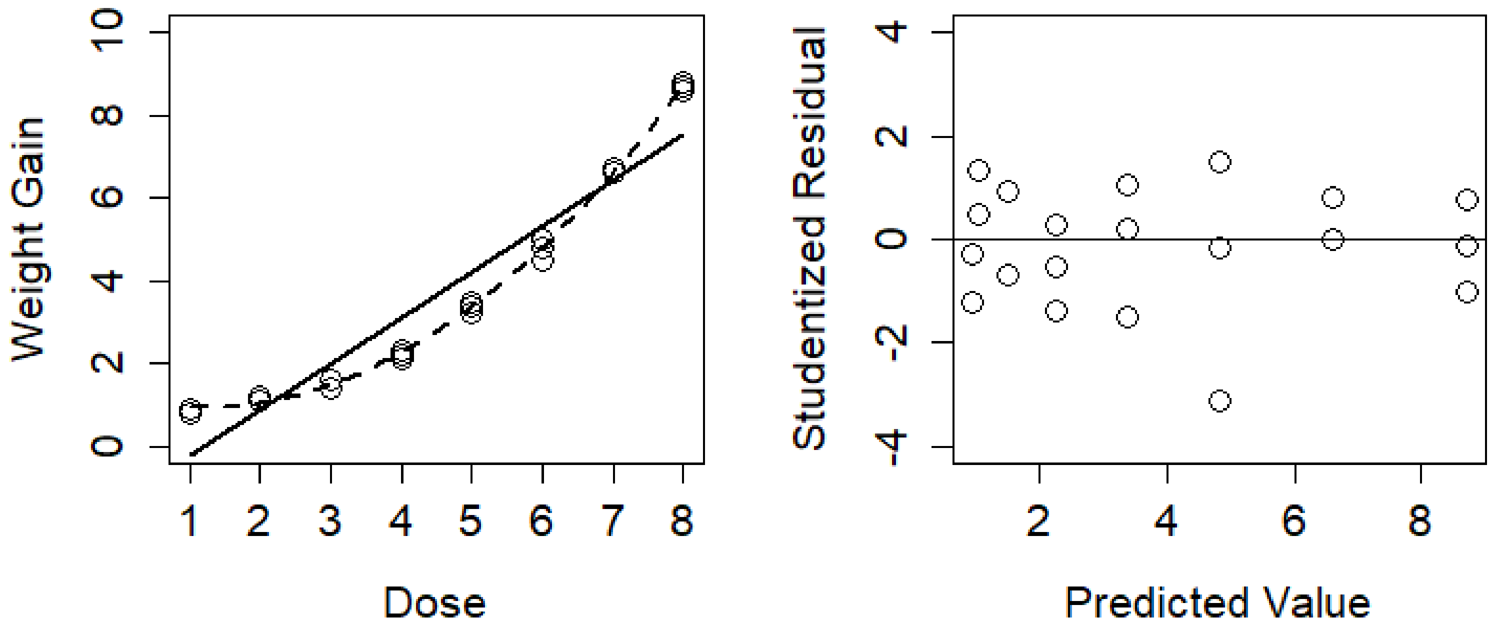
No. There is an obvious pattern in the residuals.

Is there a significant linear association between drug dose and weight loss?

Yes. From the previous slide, on average, there is a 1.11 unit increase per unit increase in dose (t=15.44 (or F = 238.27), p<0.0001).

Example: Quadratic vs Linear Model

What if we fit a polynomial model of order 2?




```
DATA wtgain;
    set wtgain;
    dosesq = dose**2; /* "**" indicates power, in this case squared */

    /* create dummy variables */
    dose1 = (dose=1);
    dose2 = (dose=2);
    ...
    dose8 = (dose=8);
RUN;

PROC REG DATA=wtgain;
    MODEL wgtgain = dose dosesq / covb;
RUN;
```

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	2	**169.70004**	84.85002	5247.78	<.0001
Error	21	0.33954	**0.01617**		
Corrected Total	23	170.03958			

MSE reduced dramatically (SSE reduced, DF decreased)

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	**1.15536**	0.10242	11.28	<.0001
dose	1	**-0.39028**	0.05222	-7.47	<.0001
dosesq	1	0.16687	0.00566	29.46	<.0001

Note that b_0 and b_{dose} estimates changed when b_{dosesq} was added to the model.

Covariance of Estimates			
Variable	**Intercept**	**dose**	**dosesq**
Intercept	0.0104904359	-0.004908369	0.0004812127
dose	-0.004908369	0.0027268717	-0.000288728
dosesq	0.0004812127	-0.000288728	0.0000320808

Example: Quadratic Model cont.

Questions about the association between drug dose and weight loss for the quadratic model:

- Is the overall regression significant? That is, is more of the variation in Y explained by the second-order model than by ignoring **X** completely (and just using $\bar{Y}$).

<u>Overall F Test</u>

F = 5247.78 $p < 0.0001$

- Does the second-order model provide significantly more predictive power than the straight-line model does?

<u>Partial F test or t statistic for the beta coefficient for the quadratic term</u>

t = 29.46 $p < 0.0001$

- Given that a second-order model is more appropriate than a straight-line model, should we add higher order terms to the second-order model?

  It is possible adding higher order terms may be beneficial, but we must balance this with our consideration of identifying a parsimonious model. Further exploration will be needed.

C. Estimate of $\sigma^2_{Y|X}$ and Lack of Fit

Given that a second-order model is more appropriate than a straight-line model, should we add higher order terms to the second-order model?

Recall that we use **mean square error** (***MS_{Error}***) to estimate $\sigma^2_{Y|X}$. This estimate is given by the formula:

$$\hat{\sigma}^2_{Y|X} = \frac{\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2}{n-2} = \frac{SS_{Error}}{n-2} = MS_{Error}$$

The MS_{Error} will only provide an unbiased estimate of the error variance when the hypothesized model is correct (in this case, if a straight-line model is appropriate), otherwise the MS_{Error} will estimate a quantity larger than the error variance: $\hat{\sigma}^2_{Y|X} > \sigma^2_{Y|X}$.

- If the model is incorrect, then two factors contribute to the inflation of the SSE. The true variability in *Y* (the "***pure error***") and the error due to fitting an incorrect model (the "***lack of fit***" error)
- With replicate observations (i.e., multiple observations with the same values of the predictor(s)), we can test for lack of fit of the assumed model by obtaining an estimate of $\sigma^2_{Y|X}$ that does not assume the correctness of the straight-line model (it is a model-free estimate of the residual variance or the "pure error").

Calculation of sums of squares due to *pure error*:

Dose (X)	Weight Gain (Y)			$\bar{Y}_x$	SS_{PE} $\sum_m (Y_{mx} - \bar{Y}_x)^2$	df
1	0.9	0.9	0.8	0.866667	.006667	2
2	1.1	1.1	1.2	1.133333	.006667	2
3	1.6	1.6	1.4	1.533333	.026667	2
4	2.3	2.1	2.2	2.200000	.020000	2
5	3.5	3.4	3.2	3.366667	.046667	2
6	5.0	4.5	4.8	4.766667	.126667	2
7	6.6	6.7	6.7	6.666667	.006667	2
8	8.7	8.6	8.8	8.700000	.020000	2
					Σ=0.26	**Σ=16**
				MS_{PE} = 0.26/16 = 0.01625		

$\hat{\sigma}^2_{Y|X}$

This estimate $\hat{\sigma}^2_{Y|X}$ is not model dependent. It is the "pure error".

So, to test the linear trend using the "pure error":

$$t = \frac{\hat{\beta}_{dose}}{\sqrt{\frac{MSE(pure)}{MSE(pure+LOF)} \times (SE(\hat{\beta}_{dose}))^2}} = \frac{1.11151}{\sqrt{\frac{0.01625}{0.65331} \times (0.072012)^2}} = \frac{1.11151}{\sqrt{0.02487 \times (0.072012)^2}} = 97.87$$

Model with just dose

$$F = t^2 = 9578.54$$

Previous t=15.44.
Variance was 40 times higher due to lack of fit error.

Estimating the "*pure error*"

You can also obtain the "pure error" by fitting a "saturated model" – a model using a dummy code for each level *k* of X (or *k* -1 dummy codes if fitting an intercept as below)

```
PROC REG DATA=wtgain;
    MODEL wgtgain = dose2 dose3 dose4 dose5 dose6 dose7 dose8;
RUN;
```

Dose Group	Variable Coding for Model						
	dose2	dose3	dose4	dose5	dose6	dose7	dose8
1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0
3	0	1	0	0	0	0	0
4	0	0	1	0	0	0	0
5	0	0	0	1	0	0	0
6	0	0	0	0	1	0	0
7	0	0	0	0	0	1	0
8	0	0	0	0	0	0	1

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	7	169.77958	24.25423	1492.57	<.0001
Error	16	0.26000	0.01625		
Corrected Total	23	170.03958			

Pure Error

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	0.86667	0.07360	11.78	<.0001
dose2	1	0.26667	0.10408	2.56	0.0209
dose3	1	0.66667	0.10408	6.41	<.0001
dose4	1	1.33333	0.10408	12.81	<.0001
dose5	1	2.50000	0.10408	24.02	<.0001
dose6	1	3.90000	0.10408	37.47	<.0001
dose7	1	5.80000	0.10408	55.72	<.0001
dose8	1	7.83333	0.10408	75.26	<.0001

The intercept is the predicted weight gain when dose2 – dose8 are zero, which occurs for dose group 1 (the reference group). Thus the intercept is the predicted weight gain (which is equal to the observed weight gain) for dose group 1.

Lack-of-Fit

The difference in the Regression Sum of Squares between the lower-order model being considered and the full model containing all higher-order terms is the ***lack of fit sum of squares***.

The ***lack-of-fit test statistic*** is a partial F test for testing the addition of the higher-order terms (up to the highest order) to the polynomial model.

$$F = \frac{[SS_{error}(reduced) - SS_{error}(full)]/k}{MS_{error}(full)} = \frac{[SS_{\text{model}}(full) - SS_{\text{model}}(reduced)]/k}{MS_{error}(full)} \sim F_{k,n-p-k-1}$$

Note: the <u>error</u> sum of squares for the highest-order polynomial model is equivalent to the <u>error</u> sum of squares for a model including a dummy variable for each dose level without an intercept (cell means model) or leaving out a reference group if an intercept is included (reference cell/group model).

The <u>model</u> sum of squares for the highest order polynomial model and reference group model will also be the same, but not the same as the cell means model (since the cell means model uses the "uncorrected sum of squares").

Testing Lack of Fit

$$F = \frac{[SS_{error}(reduced) - SS_{error}(full)]/k}{MS_{error}(full)} = \frac{[SS_{\text{model}}(full) - SS_{\text{model}}(reduced)]/k}{MS_{error}(full)} \sim F_{k,n-p-k-1}$$

Lack-of-Fit Test for the straight-line model:

H$_0$: $\beta_{quad} = \beta_{cubic} = \beta_{quartic} = \ldots = \beta_{septic} = 0$

H$_0$: $\beta_{X^2} = \beta_{X^3} = \ldots = \beta_{X^7} = 0$

Model with just dose

$$\text{F} = \frac{(14.372990\text{-}0.2600)/6}{0.01625} = \frac{(169.77958\text{-}155.66669)/6}{0.01625} = \frac{14.1129/6}{0.01625} = 144.75 \sim F_{6,16}; p < 0.0001$$

Saturated model

Lack-of-Fit Test for the quadratic model:

H$_0$: $\beta_{cubic} = \beta_{quartic} = \ldots = \beta_{septic} = 0$

H$_0$: $\beta_{X^3} = \ldots = \beta_{X^7} = 0$

Model with dose & dose squared

$$\text{F} = \frac{(169.77958\text{-}169.70004)/5}{0.01625} = \frac{0.07954/5}{0.01625} = 0.979 \sim F_{5,16}; p = 0.460$$

Saturated model

Fitting and Testing Higher-Order Models

How large an order of polynomial model to consider depends on the problem being studied and the amount and type of data being collected.

A large number of well-placed predictor values and a small error variance are needed to obtain reliable fits for models of higher than order 3 (cubic).

Fitting polynomial models or orders higher than three usually leads to models that are neither always decreasing nor always increasing.

- Substantial theoretical and/or empirical evidence should exist to support the employment of such complicated non-monotonic models.

The quantity of data directly limits the maximum order of a polynomial that may be fit.

- Generally, the maximum-order polynomial that may be fit is one less than the number of distinct X-values (a polynomial curve of order d-1 can pass exactly through d distinct X-values).

Collinearity problems can arise in polynomial models

- Centering the predictors can help remedy such problems for the second-order polynomial model

- Orthogonal polynomials can also be used (see supplemental Lecture S1 on Contrasts)

D. More on Polynomial Regression: Hierarchical Models

Consider the polynomial model of order 2 (the quadratic model):

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$$

Suppose we fit this model and the coefficient for *x*, b_1, is not significant but the coefficient for x^2, b_2, is significant. If we then removed the *x* term, then our reduced model becomes:

$$y = \beta_0 + \beta_2 x^2 + \epsilon$$

But suppose we then made a location change $x \rightarrow x+z$, where *z* is a constant. Then the model would become:

$$y = \beta_0 + \beta_2 x^2 + 2\beta_2 xz + \beta_2 z^2 + \epsilon$$

The 1st order *x* term has now reappeared so our model has effectively changed.

- Location changes should <u>not</u> make any important change to the model but in this case an additional term has been added.
- This is one reason why we should not remove lower order terms in the presence of higher order terms -- we would not want the conclusion to depend on the choice of location.

Hierarchical Models cont.

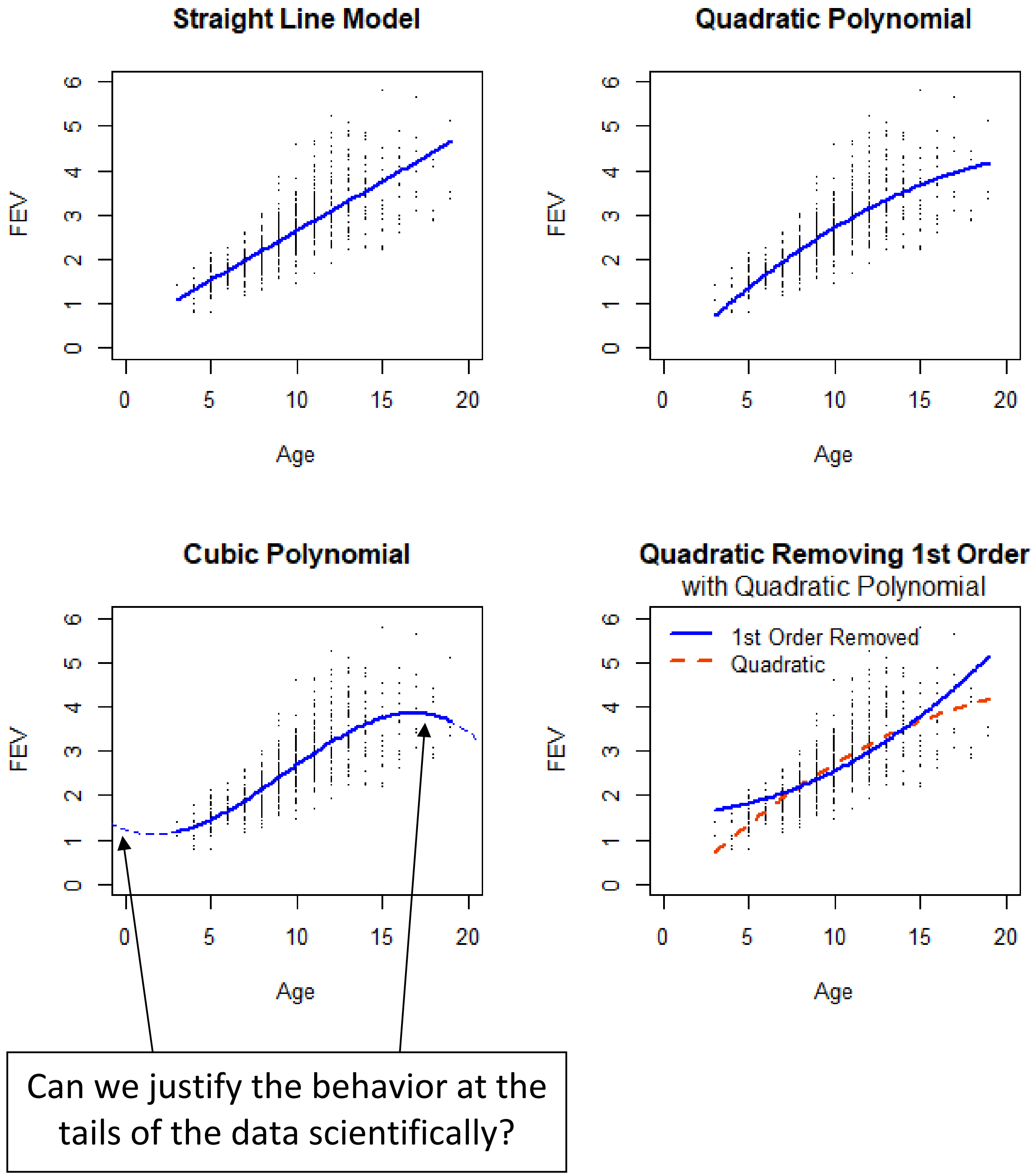
In addition, removal of the 1st order term corresponds to the hypothesis that the predicted response is symmetric about x=0 and has an optimum (minimum or maximum) at x=0.

- Often this hypothesis is not meaningful and should not be considered.
- Only when this hypothesis is scientifically justifiable should we consider removing the lower order term.

Main Point: In general, you want to maintain a hierarchical model

- Keep lower order terms in the model
- Note: A similar argument can be made about removing β_0 from a model.

Polynomial Models for FEV data



```
/* straight line model */
PROC REG DATA=fev;
    MODEL fev = age;
RUN;

/* quadratic model */
PROC REG DATA=fev;
    MODEL fev = age agesq;
RUN;

/* cubic model */
PROC REG DATA=fev;
    MODEL fev = age agesq agecu;
RUN;

/* quartic model */
PROC REG DATA=fev;
    MODEL fev = age agesq agecu agequ;
RUN;

/* quadratic model with 1st order polynomial removed */
PROC REG DATA=fev;
    MODEL fev = agesq;
RUN;
```

Linear Model:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	280.91916	280.91916	872.18	<.0001
Error	652	210.00068	0.32209		
Corrected Total	653	490.91984			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.43165	0.07790	5.54	<.0001
age	1	**0.22204**	**0.00752**	**29.53**	**<.0001**

Quadratic Model:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	286.70559	143.35280	456.98	<.0001
Error	651	204.21424	0.31369		
Corrected Total	653	490.91984			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.36036	0.19979	-1.80	0.0717
age	1	0.38571	0.03882	9.94	<.0001
agesq	1	**-0.00776**	**0.00181**	**-4.29**	**<.0001**

Cubic Model:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	290.87948	96.95983	315.06	<.0001
Error	650	200.04035	0.30775		
Corrected Total	653	490.91984			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.23226	0.47558	2.59	0.0098
age	1	-0.13324	0.14607	-0.91	0.3620
agesq	1	0.04387	0.01413	3.10	0.0020
agecu	1	**-0.00159**	**0.00043051**	**-3.68**	**0.0002**

Quartic Model:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	291.69429	72.92357	237.56	<.0001
Error	649	199.22555	0.30697		
Corrected Total	653	490.91984			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.82562	1.08723	2.60	0.0096
age	1	-0.85652	0.46730	-1.83	0.0673
agesq	1	0.15722	0.07099	2.21	0.0271
agecu	1	-0.00894	0.00453	-1.97	0.0491
agequ	1	**0.00016800**	**0.00010312**	**1.63**	**0.1038**

Quadratic Model (removing first order polynomial):

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	255.74206	255.74206	709.01	<.0001
Error	652	235.17777	0.36070		
Corrected Total	653	490.91984			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.57789	0.04618	34.17	<.0001
agesq	1	0.00986	0.00037048	26.63	<.0001

Best Model?

Cubic model, statistically.

But does this model make sense scientifically (decreasing FEV after age 17)?

It is very difficult to interpret these coefficients. When fitting polynomial models, it is best to plot the regression equation!

THE REMAINING LECTURE NOTES ARE FYI ONLY

They will not be on the homework or exams for 6611.

E. Other Remedies for Non-Linearity

Sometimes we find that a low-order polynomial provides a poor fit to the data, and increasing the order of the polynomial or transforming *X* or *Y* doesn't help (the residual plot still exhibits some structure).

One flexible approach is the use of **spline functions** to perform **piecewise polynomial fitting**:

- Divide the range of *X* into segments.
- Fit an appropriate curve of order *k* in each segment.

The piecewise linear spline (*k*=1) with a single knot (change point between segments) is given by:

$$E(Y) = \beta_{00} + \beta_{01}X + \beta_{10}(X-T)_+^0 + \beta_{11}(X-T)_+^1$$

$$(X-T)_+ = \begin{cases} (X-T) & \text{if } X-T > 0 \\ 0 & \text{if } X-T \le 0 \end{cases}$$

If X ≤ T, the straight-line model is

$$E(Y) = \beta_{00} + \beta_{01}X$$

And if X > T then the straight-line model is:

$$E(Y) = \beta_{00} + \beta_{01}X + \beta_{10} + \beta_{11}(X-T)$$

A smoother function would result if we required the regression function to be continuous at the knot. This can be accomplished by deleting the $\beta_{10}(X-T)_+^0$ term from the model above:

$$E(Y) = \beta_{00} + \beta_{01}X + \beta_{11}(X-T)_+^1$$

If X ≤ T, the straight-line model is

$$E(Y) = \beta_{00} + \beta_{01}X$$

And if X > T then the straight-line model is:

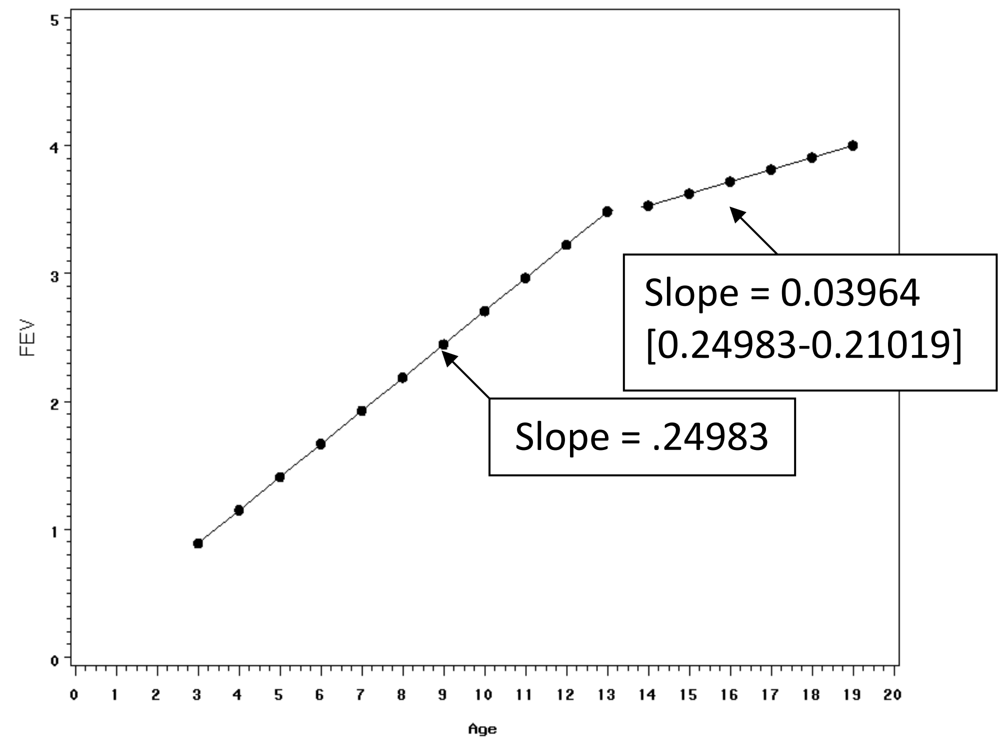
$$\begin{aligned} E(Y) &= \beta_{00} + \beta_{01}X + \beta_{11}(X-T) \\ &= (\beta_{00} - \beta_{11}T) + (\beta_{01} + \beta_{11})X \end{aligned}$$

- We usually assume that the positions of the knots are known.
  - If they are parameters to be estimated, the resulting problem is a nonlinear regression problem.
- A **cubic spline** is a piecewise polynomial of order 3 (which is usually adequate for most practical problems).
- A potential disadvantage of cubic splines is that the $\mathbf{X}^T\mathbf{X}$ matrix can become ill-conditioned if there are a large number of knots.
  - This problem can be overcome by using a different representation of the spline called the **cubic B-spline**.

Example: FEV Data (*Discontinuity at the Knot*)

```
*** Data Step Code for fitting piecewise polynomial models ***;
*** Allowing for a knot at age=14 ***;
data fev;
    set fev;

    I14 = (age ge 14);
    age14 = age-14;
    age14_I14 = I14*age14;

    LABEL   I14      = 'I(age>=14)'
            age14    = '(Age-14)'
            age14_I14 = '(Age-14)*I(age>=14)';
run;

PROC REG DATA=fev;
    MODEL fev = age I14 age14_I14;
    OUTPUT out = pred2 predicted=p;
RUN;

PROC SORT DATA=pred2;
    BY age;
RUN;

PROC GPLOT DATA=pred2;
    plot p*age / VAXIS=axis1 HAXIS=axis2;
    SYMBOL INTERPOL=join VALUE=dot COLOR=black;
RUN;
```

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	290.84721	96.94907	314.97	<.0001
Error	650	200.07262	0.30780		
Corrected Total	653	490.91984			

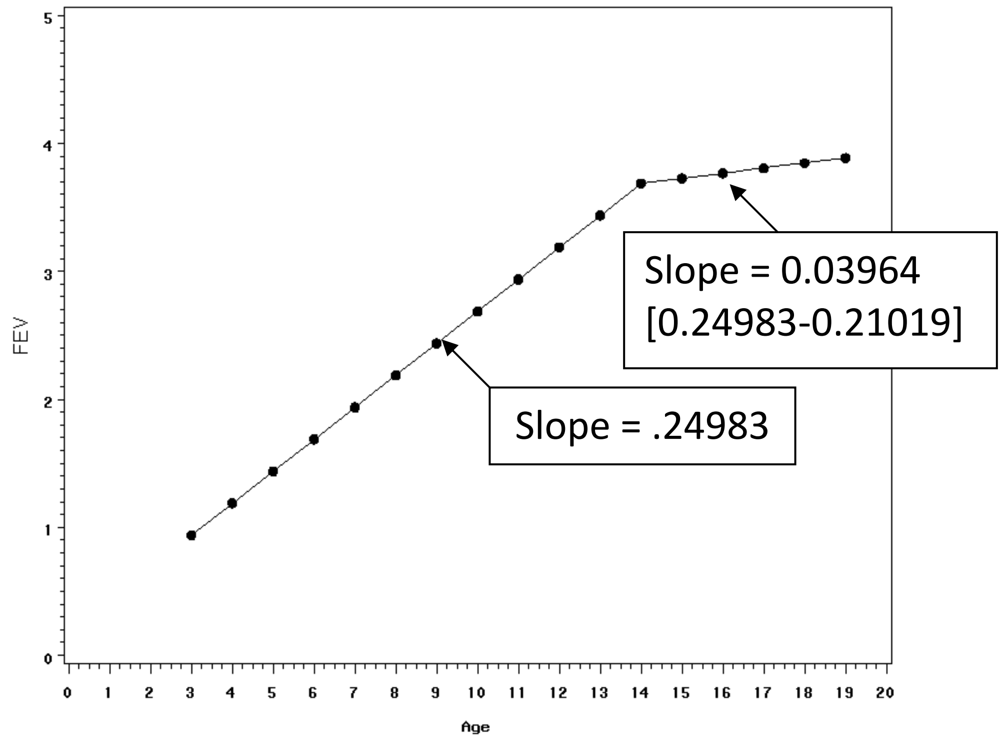


	Parameter Estimates						
	Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
β_{00}	**Intercept**	Intercept	1	0.11439	0.09638	1.19	0.2357
β_{01}	**age**		1	0.25916	0.01014	25.55	<.0001
β_{10}	**I14**	I(age>=14)	1	-0.21310	0.10596	-2.01	0.0447
β_{11}	**age14_I14**	(Age-14)*I(age>=14)	1	-0.16499	0.04551	-3.63	0.0003

***Continuity at the Knot*:**

```
PROC REG DATA=fev;
    MODEL fev = age age14_I14;
    OUTPUT out = pred3 predicted=p;
RUN;

PROC SORT DATA=pred3;
    BY age;
RUN;

PROC GPLOT DATA=pred3;
    PLOT p*age / VAXIS=axis1 HAXIS=axis2;
    SYMBOL INTERPOL=join VALUE=dot COLOR=black;
RUN;
```

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	2	289.60233	144.80116	468.24	<.0001
Error	651	201.31751	0.30924		
Corrected Total	653	490.91984			



	Parameter Estimates						
	Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
β_{00}	**Intercept**	Intercept	1	0.19034	0.08888	2.14	0.0326
β_{01}	**age**		1	0.24983	0.00904	27.63	<.0001
β_{11}	**age14_I14**	(Age-14)*I(age>=14)	1	-0.21019	0.03967	-5.30	<.0001

Nonparametric Methods

Nonparametric regression models are also available, including **Kernel** regression and **Loess,** both of which use data from a "neighborhood" around specific locations to estimate the regression line.

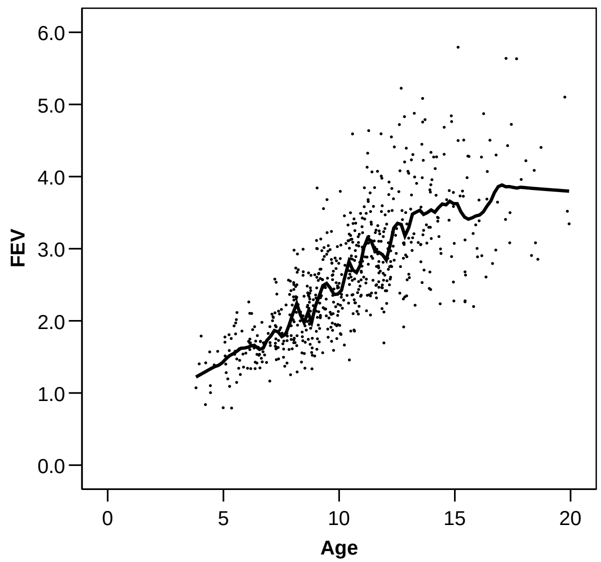
LOESS (locally weighted polynomial regression)

- At each point in the data set a low-degree polynomial is fit to a subset of the data, with explanatory variable values near the point whose response is being estimated.
- The polynomial is fit using weighted least squares, giving more weight to points near the point whose response is being estimated and less weight to points further away.
- The value of the regression function for the point is then obtained by evaluating the local polynomial using the explanatory variable values for that data point.
- The LOESS fit is complete after regression function values have been computed for each of the ***n*** data points.
- Many of the details of this method, such as the size of the neighborhood, degree of the polynomial model and the weights, are flexible and can be chosen by the analyst.

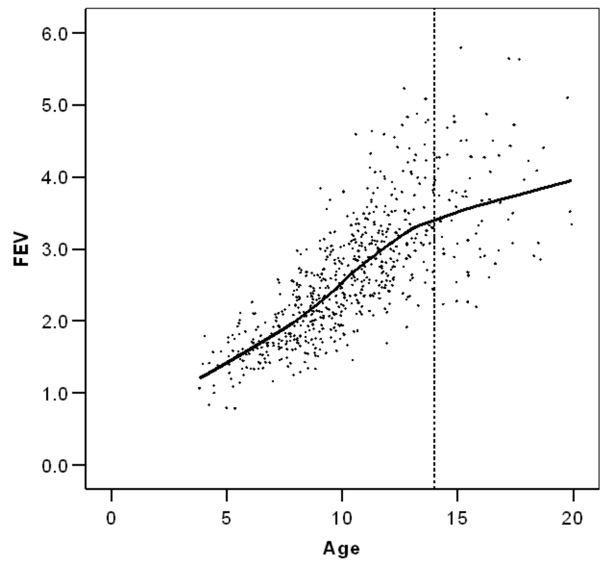
```
*** LOESS REGRESSION ***;
PROC LOESS DATA=fev;
 MODEL fev=age / smooth=.99;
 ODS OUTPUT outputstatistics=stats;
RUN;
```

```
PROC SORT DATA=stats;
 BY age;

PROC GPLOT DATA=stats;
  PLOT (depvar pred)*age /overlay;
  symbol1 c=black i=rl value=dot width=2;
  symbol2 c=red i=join value=none width=2;
RUN;
```

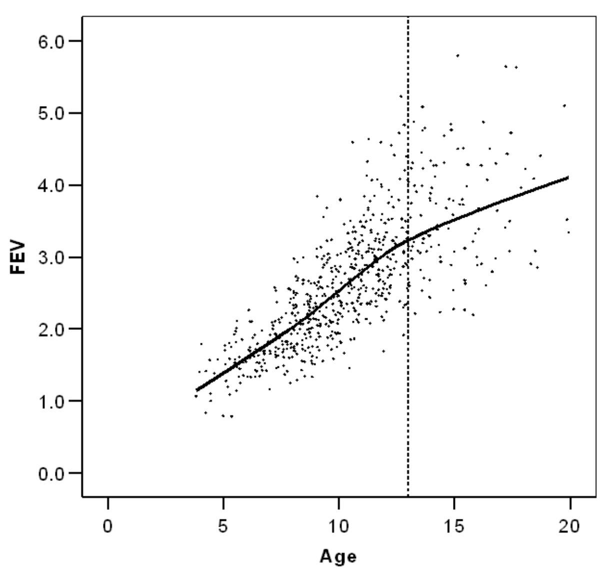
Loess: 5% of data points



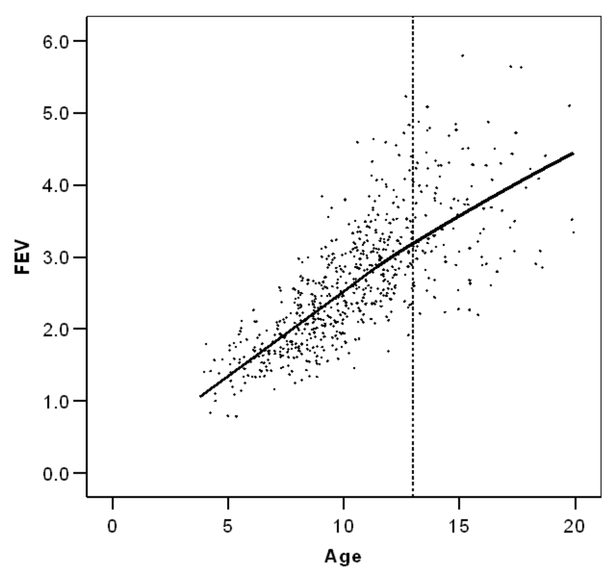
Loess: 50% of data points



Loess: 75% of data points



Loess: 99% of data points



PROC GPLOT with 99% of data points (black is linear fit, red is loess fit):