Chapter 8 Polynomial Regression

used when

1) True curvilinear response is a polynomial function
2) The curvilinear responce is unknown but a polynomial is a good approximation

* Warning: This can have disasterous results in extrapolation

Second Order model w/ one predictor $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$

NOTE: we can use $x_i = X_i - \bar{X}$ to reduce multicollinearity

$E(Y_i) = \beta_0 + \beta_1 x + \beta_2 x^2$

$\beta_2 = \beta_{11}$

Third Order Model w/ one predictor $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_i$

NOTE: we can use $x_i = X_i - \bar{X}$ to reduce multicollinearity

$E(Y_i) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

$\beta_2 = \beta_{11}$, $\beta_3 = \beta_{111}$

CAUTION: We try not to go beyond the third order in polynomial regression, instead we employ other models

Second Order model w/ two predictors $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \epsilon$

NOTE: we can use $x_{i1} = X_{i1} - \bar{X}_1$; $x_{i2} = X_{i2} - \bar{X}_2$

$E(Y_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$

Hierarchical Approach to Fitting - We don't want to include higher order terms w/o including the first order predictors

Sometimes we fit the third order and, then, eliminate predictors as we go.

SAS

SAME AS Regular Regression

GLM:

```
proc glm;
   model y = x1 x1*x1 x1*x1*x1 x2 x3 x1*x2 /solution;
run;
```

REG

```
data PolyRegData;
   set origDataName;
   x1sq =
   x1cu =
   x1x2 =
run;
proc reg;
   model y = x1 x1sq xcu x2 x3 x1x2;
run;
```

NOTE • REG needs to be done in a data step

• OUTPUT IS the same

oops!

34

Interaction Regression Models $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i$

$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$

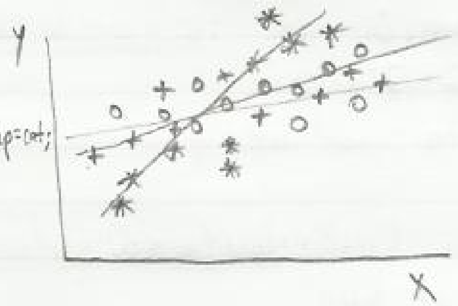
Interpretation - The mean response to a unit change in x_1 is $\beta_1 + \beta_3 x_2$ all else constant, of course

- The mean response to a unit change in x_2 is $\beta_2 + \beta_3 x_1$ all else constant

How to tell if we need an interaction term

SAS

```
proc sgscatter;
  plot y*x / reg group=cat;
run;
```



G_1 o

G_2 +

G_3 *

Note: If we were to draw three best fit lines they would have different slopes

NOTE: Again we can use $x_{ik} = X_{ik} - \overline{X}_k$ to reduce multicollinearity

How to test $\beta_3 = \beta_4 = \beta_5 = 0$

Sas just add test statement to proc reg

```
proc reg;
  model = x1 x2 x1^2 x2^2 x1x2;
  test X3 = X4 = X5 = 0;
run;
```

Qualitative Predictors $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$ $x_1 = \{ {1 \atop 0} {G1 \atop G2} \}$ $x_2 = \{ {1 \atop 0} {G2 \atop G1} \}$

categorical

* Regression lines will be // with different slopes dependent on the categorical predictors

$E(Y) = (\beta_0 + \beta_1) + \beta_3 x_{i3}$ for G_1

$E(Y) = (\beta_0 + \beta_2) + \beta_3 x_{i3}$ for G_2

* Type III tests are used to see whether or not we can drop categorical predictors from the model
* We can do this in proc reg if we create 0/1 dummies

$$z_j = \begin{cases} 1 & \text{category} = j \\ 0 & \text{ow} \end{cases}$$

Collapsing levels

SAS

```
proc reg;
  model y = x1 dummy1 dummy2;
  test dummy1 - dummy2 = 0;
run;
proc glm
  class cat;
  model y = x1 cat;
  contrast 'TITLE' cat 0 1 -1;    Can I collapse Group 2 and Group 1
run;
```

Interactions between Quatitative and Categorical variables

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i$ where x_2 is categorical

Now, we notice the slope will change when the category changes thus, they are no longer parallel.