Polynomial Regression:

We mentioned the possibility of fitting curvilinear functions through the use of multiple regression –

Based on one independent variable, we could consider a model of the form:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \cdots + \beta_{p-1} X_{i1}^{p-1} + \epsilon_i$$

Here, we have a $p-1$ order polynomial and all the various testing procedures for the multiple regression model apply (as your homework indicates)

Some points to make:

1) What is the highest order polynomial that could be fit?

Suppose we have repeats at certain values of X;

if there are c distinct values of X, then a polynomial of degree $c-1$ fits the means perfectly;

anything higher leads to a singular $\boldsymbol{X}'\boldsymbol{X}$

Usually, anything more than 3 is superfluous and would represent overfitting.

Stick with linear, quadratic, and at most cubic.

2) For numerical considerations, usually use $X - \bar{X}$ instead of the original X;

this is an issue of tolerance (i.e., one minus the squared multiple correlation between a particular independent variable and the rest)

3) Be careful about extrapolating beyond the range of X values one has

4) If there are repeats, then one can get an $MSPE$ estimate, using $n - c$ degrees of freedom;

this is the same as the MSE for a model of order $c - 1$

So, if one has a regression model of order k (less than $c-1$), then

$SSLF$ is equal to SSE for the k^{th}-order model minus the $SSPE$ obtained from the order $c-1$ model

this $SSLF$ is attributable to models of order $k+1$ to $c-1$ and denoted

$SSR(X^{k+1},\ldots,X^{c-1}|X_1,\ldots,X^k)$ with $c-1-k$ degrees-of-freedom

There are $n-k-1$ df for SSE (i.e., $(n-p)$ where $p=k+1$) and $n-c$ df for $SSPE$ (i.e., $(n-p)$ where $p=c$;

thus, $n-k-1$ minus $n-c$ equals $c-1-k$ df

Thus, $MSLF = SSLF/(c-1-k)$, so we compare

$\frac{MSLF}{MSPE} \sim F_{c-1-k,\ n-c}$

Response surface methodology involves the extension of polynomial regression to two or more independent variables.

For example, this is a second-order model with two independent variables:

$Y_i =$

$\beta_0+\beta_1 X_{i1}+\beta_2 X_{i1}^2+\beta_3 X_{i2}+\beta_4 X_{i2}^2+\beta_5 X_{i1} X_{i2}+$
ϵ_i

$E(Y_i) =$

$\beta_0+\beta_1 X_{i1}+\beta_2 X_{i1}^2+\beta_3 X_{i2}+\beta_4 X_{i2}^2+\beta_5 X_{i1} X_{i2}$

gives the equation for a conic section.

Generally, need three dimensions to plot: Y, X_1, and X_2

In the X_1 and X_2 plane, constant values for $E(Y_i)$ might, for example, lead to "concentric" ellipses