Regression Models for Quantitative and Qualitative Predictors: An Overview

- Polynomial regression models
- Interaction regression models
- Qualitative predictors
- Indicator variables
- Modeling interactions between quantitative and qualitative predictors

Types of Polynomial Regression Models

- Using a polynomial in X to predict Y: accommodate nonlinearity
  - Multicollinearity.
  - Regression coefficients for lower-order terms vary wildly as higher-order terms are added.
- To reduce Multicollinearity, one can use centered predictors.

$$X_{ik}^* = X_{ik} - \bar{X}_k, \quad k = 1, \ldots, p-1, \quad i = 1, \ldots, n.$$

  - Then fit regression in the normal way:

$$Y_i = \beta_0 + \beta_1 X_{i1}^* + \beta_2 {X_{i1}^*}^2 + \epsilon_i.$$

  - Guideline: If X^k is to be included in the model, so must all lower-order terms regardless of their significance.

Polynomial Regression Models: Remarks

- Polynomial regression models can be useful when
  - the true response surface is a polynomial function
  - the true response surface is unknown/complex but a polynomial function offers a good approximation
- It is very rare that one needs to fit high order models (> 4), as including many higher order terms (main and interaction effects) in the model increases the risk of:
  - interpolating the noise rather than estimating the underlying trend
  - extrapolation beyond the range of the data
- Implementation: Conduct parameter estimation, estimation of mean response, prediction of new observation, and model diagnostics and assessment as in multiple linear regression.

Orthogonal Polynomials

- Orthogonal polynomials are 100% effective at eliminating multicollineary
  - Available in R as `y~poly(x,order)`
  - The regression coefficient on each successive polynomial term can be calculated independently of the previous ones
  - The previous (i.e., lower-order) regression coefficients don't change when the next term is added to the model. Why?
- The SS attributed to each polynomial term is independently calculated and represents the amount by which the SSR is increased by passage from an equation of lower degree.

Orthogonal Polynomials

- For example: let $\phi_j(X)$ denote the orthogonal polynomial of degree j.

$$
\begin{aligned}
\phi_0(X) &= 1 \\
\phi_1(X) &= X - \bar{X} \\
&\vdots \\
\phi_j(X) &= X^j - (\alpha_{0,j} + \alpha_{1,j}\phi_1(X) + \cdots + \alpha_{j-1,j}\phi_{j-1}(X))
\end{aligned}
$$

  where $\alpha_{i,j}$ are the parameters of the regression equation when X^j is regressed on $\phi_1(X), \ldots, \phi_{j-1}(X)$.

- The geometric interpretation mentioned previously provides intuition that these polynomials are orthogonal because the residuals from the jth orthogonalizing regression are orthogonal to the jth response (X^j)

- Gram-Schmidt procedure

  - GS-procedure does not alter the prediction essentially, why?

- Other choice of orthogonal basis functions

Interaction Effects

- A regression model is additive if the response surface is in the form of

$$\mathbb{E}(Y) = \sum_{j=1}^{p} f_j(X_j)$$

  - Example: Additive model.

$$\mathbb{E}(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2.$$

  - Counter example: Model contains an interaction effect.

$$\mathbb{E}(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2.$$

- Interactions are used when the effect of one predictor variable depends on the level(s) of the other predictor variable(s).

Interaction Effects

- Consider $Y = \beta_0 + \beta_1 X_1 + \beta_2 I(Female) + \beta_3 X_1 I(Female)$.
- For males, regression line is $Y = \beta_0 + \beta_1 X_1$
- For females, $Y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1$
- The two genders have wholly different regression lines
  - Different intercepts
  - Different slopes
  - Males are the baseline
  - The shift for females is modeled as a change from the male level
  - The X_1 slope for females is modeled as a change from the male slope
- Same notion for $p - 1 > 2$ (fitting hyperplanes)

Interaction Effects

- Consider $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$. The partial derivatives are
  - $\partial / \partial X_1 = \beta_1$
  - $\partial / \partial X_2 = \beta_2$
- Now consider $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$.
  - $\partial / \partial X_1 =$
  - $\partial / \partial X_2 =$
- Fully general second-order polynomial:

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

Implementation of Interaction Regression Models

- Adding interaction terms to a regression model may induce multicollinearity (e.g. X_1 is related to X_1X_2).
  - Multi-way interactions are allowed, e.g., $X_1X_2X_3$. With $p-1$ predictors, the $(p-1)$-interaction is generally not allowed as its SSR is SSE, i.e. the model is overparameterized and has no residual degrees of freedom.
- When the # of predictor variables is larger, the # of possible interaction terms can be huge (e.g. 2-, 3-, 4-way interactions).
  - RULE: If a predictor variable participates in an interaction, all the terms lower-order terms in this predictor variable are typically retained in the model, regardless of their significance. This is also known as the hierarchical approach to model fitting.

Qualitative Predictor Variables

- A qualitative variable records into which of several categories a subject falls.
  - As a result, arithmetic operations (e.g. mean, difference) do not make sense for the categories.
  - Examples:
    - ⋆ Gender.
    - ⋆ Soil type.
- Outline:
  - Two-category (binary) predictor variables.
  - Multi-category predictor variables.

Binary Predictor Variables

- Example A: Automobile mileage.

$$Y = \text{mileage (miles per gallon)}$$
$$X_1 = \text{weight}$$
$$X_2 = \text{I(hybrid)}$$

- Question A.1: Is there a relationship between X_1 and Y?
- Consider the model (A.1):
- The test of interest is H_0 :

Binary Predictor Variables

- Question A.2: For a given weight X_1, is there a constant (additive) difference in mileage between hybrid and standard cars?
- Consider the model (A.2):
  - For a standard car $X_2 = 0$, $\mathbb{E}(Y) =$
  - For a hybrid car $X_2 = 1$, $\mathbb{E}(Y) =$
- The test of interest is H_0 :

Binary Predictor Variables

- Question A.3: For any given weight X_1, is there any difference in the effect on mileage of hybrid versus standard cars?
- Consider the model (A.3):
  - For a standard car $X_2 = 0$, $\mathbb{E}(Y) =$
  - For a hybrid car $X_2 = 1$, $\mathbb{E}(Y) =$
- An overall test of interest is H_0 :
- If the overall test is significant, then for testing interactions we test H_0 :
- Why not just fit separate simple linear regression models for hybrid and standard cars?

Example: `drug`

- A study was conducted to assess the effect of drug dose levels on thyroid activity for two different drugs.
- The data consist of the dose levels (X) and measurements of thyroid activities (Y).
- A question of interest is whether the relation between dose and activity is similar for the two drugs.

Example: drug

- The original data have the following format:

Drug A		Drug B	
X_1	Y_1	X_2	Y_2
X_{11}	Y_{11}	X_{21}	Y_{21}
X_{12}	Y_{12}	X_{22}	Y_{22}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_{1n_1}	Y_{1n_1}	X_{2n_2}	Y_{2n_2}

Example: drug

Rearrange the data to the following format:

W_1	W_2	$W_3 = W_1 W_2$	Y
X_{11}	1	X_{11}	Y_{11}
X_{12}	1	X_{12}	Y_{12}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_{1n_1}	1	X_{1n_1}	Y_{1n_1}
X_{21}	0	0	Y_{21}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_{2n_2}	0	0	Y_{2n_2}

- $W_2 = I(\text{Drug A})$ and W_3 represents the interaction term
- The model is

Example: drug

- How does the multiple linear regression model relate to the simple linear regression models?
  - For drug A, the model can be conceived as

    $Y =$

  - For drug B, the model can be viewed as

    $Y =$

Example: drug

Hence, the interpretation of the slopes in the MLR is:

- β_0 is the intercept of drug B regression line β_0^B.
- β_1 is the slope of drug B regression line β_1^B.
- $\beta_2 = \beta_0^A - \beta_0^B$ is the intercept difference between drug A and B regression line.
- $\beta_3 = \beta_1^A - \beta_1^B$ is the slope difference between drug A and B regression line.

Example: drug

Various hypotheses H_0 of interest are:

- $H_0 : \beta_2 = \beta_3 = 0 | \beta_0, \beta_1$
  - We test ________________________________
  - The reduced model under H_0 is

    ____
- $H_0 : \beta_3 = 0 | \beta_0, \beta_1, \beta_2$
  - We test ________________________________
  - The reduced model under H_0 is

    ____
- $H_0 : \beta_1 = \beta_3 = 0 | \beta_0, \beta_2$
  - We test ________________________________
  - The reduced model under H_0 is

    ____

Example: drug

- From the `R` output, for drug A, the fitted simple linear regression (SLR(A)) is

$$\hat{Y} = 4.95^* + 0.564^* X$$

  and

$$MSE(A) = 0.3621, SSE(A) = 2.8967, df(A) = 8.$$

- For drug B, the fitted simple linear regression (SLR(B)) is

$$\hat{Y} = 0.691 + 1.34^* X$$

  and

$$MSE(B) = 0.292, SSE(B) = 2.338, df(B) = 8.$$

- How different are the two regression lines?

Example: drug

- From the R output, the fitted multiple linear regression is

$$\hat{Y} = 0.691 + 1.34^{*}W_1 + 4.26^{*}W_2 - 0.78^{*}W_3$$

  and $MSE = 0.327, SSE = 5.235, df = 16.$

- Note that
  - $\hat{\beta}_0 = 0.691, \hat{\beta}_1 = 1.34$ are the same as SLR(B).
  - $\hat{\beta}_0 + \hat{\beta}_2 = 0.691 + 4.26 = 4.95$ is the same as intercept SLR(A).
  - $\hat{\beta}_1 + \hat{\beta}_3 = 1.34 - 0.78 = 0.56$ is the same as slope SLR(A).
  - $SSE = 5.235 = SSE(A) + SSE(B) = 2.8967 + 2.338.$
  - df SSE = 16 = df(A) + df(B) = 8 + 8.
  - $MSE = 0.327 = \frac{1}{16}(8MSE(A) + 8MSE(B))$ is the pooled estimate of σ^2 on 16 df.

- Caution: The above bullets are only true because we fit the interaction term.

Example: drug

- Set up testing $H_0 : \beta_2 = \beta_3 = 0 | \beta_0, \beta_1$.

Full model $Y = \beta_0 + \beta_1 W_1 + \beta_2 W_2 + \beta_3 W_3 + \epsilon$

versus

Reduced model $Y = \beta_0 + \beta_1 W_1 + \epsilon$.

  - $SSE_{(F)}$ = 5.235 on df = 16; $SSE_{(R)}$ = 11.279 on df = 18.
  - Extra SS = $SSE_{(R)} - SSE_{(F)}$ on df = 2.
  - Compute

$$F^* = \frac{(SSE_{(R)} - SSE_{(F)})/(df_{(R)} - df_{(F)})}{SSE_{(F)}/df_{(F)}} = \frac{(11.279 - 5.235)/2}{5.235/16} = 9.236$$

on df = (2,16). The p-value is $P(F(2, 16) \geq 9.236) = 0.0022$.

  - There is strong evidence against H_0. That is, the two regression lines are not the same.

Example: drug

- Set up testing $H_0 : \beta_3 = 0 | \beta_0, \beta_1, \beta_2$.

$$\text{Full model } Y = \beta_0 + \beta_1 W_1 + \beta_2 W_2 + \beta_3 W_3 + \epsilon$$

versus

$$\text{Reduced model } Y = \beta_0 + \beta_1 W_1 + \beta_2 W_2 + \epsilon.$$

  - $SSE_{(F)} = 5.235$ on df $= 16$, $SSE_{(R)} = 11.188$ on df $= 17$.
  - Extra SS $= SSE_{(R)} - SSE_{(F)}$ on df $= 1$.
  - Compute

$$F^* = \frac{(11.188 - 5.235)/1}{5.235/16} = 18.194$$

    on df = (1,16). The p-value is $P(F(1,16) \geq 18.194) < 0.001$.
  - There is strong evidence against H_0. That is, the slopes of the two regression lines are not the same.

Example: drug

- Set up testing $H_0 : \beta_1 = \beta_3 = 0 | \beta_0, \beta_2$.

Full model $Y =$

versus

Reduced model $Y =$

- $SSE_{(F)} =$ ________ on df = ________ ; $SSE_{(R)} =$________ on df = ________
- Extra SS $= SSE_{(R)} - SSE_{(F)}$ on df = ________
- Compute

$$F^* =$$

on df = ________ . The p-value is ________________.

- There is strong evidence against H_0. That is,

____

Categorical Interactions

- Consider a medical study with: $X_1 = I(Female)$ and $X_2 = I(Died)$.
- The additive model is
- The predictions are:
  - $\widehat{Y}_{Male,Alive} =$
  - $\widehat{Y}_{Male,Dead} =$
  - $\widehat{Y}_{Female,Alive} =$
  - $\widehat{Y}_{Female,Dead} =$
- Thus the effect of dying is ______________________________

Categorical Interactions

- The interaction model is

- The predictions are:
  - $\widehat{Y}_{Male,Alive} =$
  - $\widehat{Y}_{Male,Dead} =$
  - $\widehat{Y}_{Female,Alive} =$
  - $\widehat{Y}_{Female,Dead} =$

- Thus the effect of dying is ________ for males and ________ for females; ________________ is the differential effect of death for females relative to males.

Multi-Category Predictor Variables

- Example B: Automobile mileage.

$$\begin{aligned} Y &= \text{mileage (miles per gallon)} \\ X_1 &= \text{weight} \\ X_2^* &= \text{SUV, sedan, or truck.} \end{aligned}$$

- Define two dummy variables X_2, X_3 as

Category	X_2	X_3
SUV	1	0
sedan	0	1
Truck	0	0

Multi-Category Predictor Variables

- Consider an additive model (B.1):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon.$$

- For an SUV $X_2 = 1, X_3 = 0$,

$$\mathbb{E}(Y) = (\beta_0 + \beta_2) + \beta_1 X_1.$$

- For a sedan $X_2 = 0, X_3 = 1$,

$$\mathbb{E}(Y) = (\beta_0 + \beta_3) + \beta_1 X_1.$$

- For a truck $X_2 = X_3 = 0$,

$$\mathbb{E}(Y) = \beta_0 + \beta_1 X_1.$$

Multi-Category Predictor Variables

- Question B.1: Is there a constant difference in mileage among the three car types?
  - The test of interest is H_0 :
- Question B.2: ____________________________________
  - The test of interest is $H_0 : [\beta_3 = 0 | \beta_1, \beta_2]$.

Multi-Category Predictor Variables

- Now consider an interaction model (B.2):

  $Y =$

  - For an SUV $X_2 = 1, X_3 = 0$,

    $\mathbb{E}(Y) =$

  - For a sedan $X_2 = 0, X_3 = 1$,

    $\mathbb{E}(Y) =$

  - For a truck $X_2 = X_3 = 0$,

    $\mathbb{E}(Y) =$

Multi-Category Predictor Variables

Interpretation of the regression coefficients

- β_0: intercept for trucks.
- β_2: difference of intercepts between SUVs and trucks.
- β_3: difference of intercepts between sedans and trucks.
- $\beta_2 - \beta_3$: difference of intercepts between SUV's and sedans.
- β_1: slope for truck weight.
- β_4: slope difference between SUVs and trucks.
- β_5: slope difference between sedans and trucks.
- $\beta_4 - \beta_5$: slope difference between SUVs and sedans.

Multi-Category Predictor Variables

- Question B.3: Is there any difference in mileage among the three car types?
  - The test of interest is H_0 :
- Question B.4: ______________________________
  - The test of interest is $H_0 : [\beta_4 = \beta_5 = 0|\beta_1, \beta_2, \beta_3]$.

Contrast Matrices

- The coding

Category	X_2	X_3
SUV	1	0
sedan	0	1
Truck	0	0

  is called a "contrast matrix" or "design matrix". (See `contrasts()`)
- This one is called the "treatment contrasts".
- Here, "Truck" is treated as the baseline, and the other vehicle effects are expressed as differences from baseline.
- Especially useful choice when there is a natural baseline, default, or control group.
- See `contr.treatment()`

Contrast Matrices

- There are other possible contrast matrices. The "sum contrasts" are:

Category	X_2	X_3
SUV	1	0
sedan	0	1
Truck	-1	-1

- This model fits
  - β_2 and β_3 are expressing deviations of the SUV and sedan means from the grand mean, μ.
  - The effect (deviation from grand mean) for Truck is understood to be $-\beta_2 - \beta_3$ since this model is based on the constraint that the sum of the three effects is zero.
  - This constraint is what releases a degree of freedom for the parameter μ

Contrasts that are virtually never used

- Helmert contrasts (adding two categories for clarity):

Category	X_2	X_3	X_4	X_5
SUV	-1	-1	-1	-1
sedan	1	-1	-1	-1
Truck	0	2	-1	-1
Motorcycle	0	0	3	-1
Bus	0	0	0	4

- Contrasts the second level with the first, the third with the average of the first two, and so on.
- Polynomial contrasts: based on orthogonal polynomials.

Multi-Category Predictor Variables: Remarks

- Terminology:
  - ANOVA: regression model with only categorical predictors
  - ANCOVA: regression model with some continuous and some categorical predictors. (Analysis of Covariance).
- You can change the baseline or "missing" factor level in the sum and treatment contrasts; see `R` help.
- For each contrast matrix (treatment and sum), can you:
  - State the proper contrast matrix for your application.
  - State what are the most important types of hypotheses, in terms of the β's.
  - Carry out and interpret the corresponding tests using the GLT strategy.