Chapter 6

Polynomial Regression and Factors

6.1 Polynomial Regression

- Model

$$E(Y \mid X = x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_d x^d$$

$$Var(Y \mid X = x) = \sigma^2$$

- Properties
  - Rarely represent physical model
  - Sometime well describe the shape of mean function

6.1 Polynomial with several predictors

- Model: e.g. Second order mean function

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

$$Var(Y \mid X = x_i) = \sigma^2$$

interaction

- Interaction

$$\mathrm{E}(Y|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$

$$\mathrm{E}(Y|X_1 = x_1 + \delta, X_2 = x_2) = \beta_0 + \beta_1 (x_1 + \delta) + \beta_2 x_2 + \beta_{11} (x_1 + \delta)^2 + \beta_{22} x_2^2 + \beta_{12} (x_1 + \delta) x_2$$

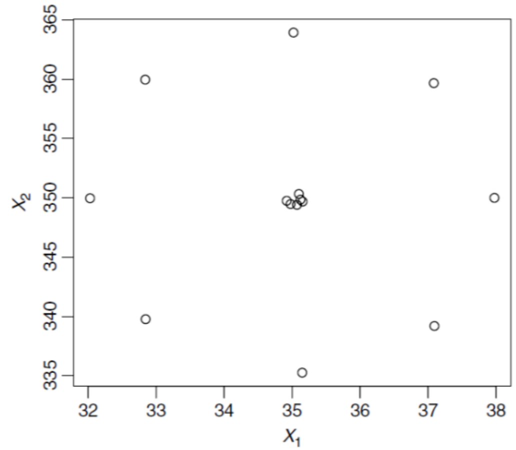
$$\longrightarrow \quad \mathrm{E}(Y|X_1 = x_1 + \delta, X_2 = x_2) - \mathrm{E}(Y|X_1 = x_1, X_2 = x_2)$$

$$= (\beta_{11}\delta^2 + \beta_1 \delta) + 2\beta_{11}\delta x_1 + \beta_{12}\delta x_2$$

Effect of a change in x_1 depends on x_2

6.1 Example

- Data
  - Y = Score for cake (Higher the better)
  - X_1 = baking time
  - X_2 = baking temperature



- Model 1 (No interaction)

$$E(Y|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$$

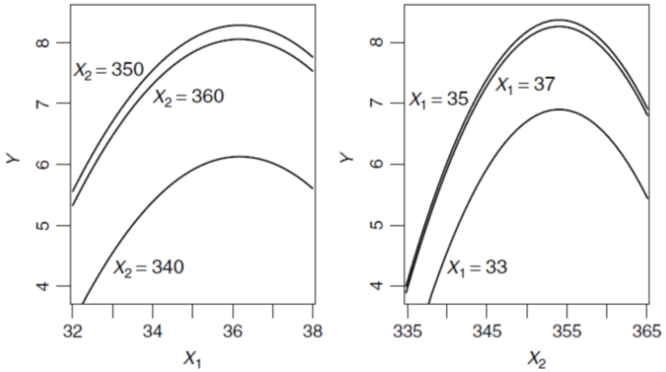
- Model 2 (Has interaction)

$$E(Y|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$

6.1 Example

- Model 1 (No interaction)

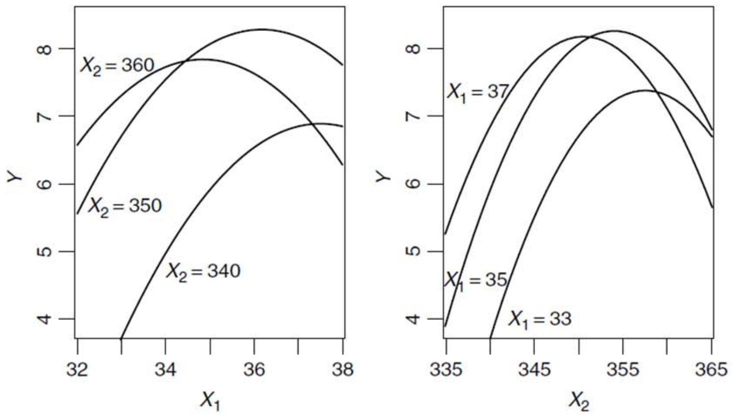
$$E(Y|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$$



The relationship between y and x_i is the same for different x_j

- Model 2 (Has interaction)

$$E(Y|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$



6.1 Example

- Model 1 (No interaction, NH)

$$E(Y|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$$

- Model 2 (Has interaction, AH)

$$E(Y|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$

- Remarks
  - We can select the best model by F-test

$$F = \frac{(RSS_{NH} - RSS_{AH})/(\mathrm{df}_{NH} - \mathrm{df}_{AH})}{RSS_{AH}/\mathrm{df}_{AH}}$$

  - We can bake the best cake by maximizing $E(Y|x_1,x_2)$ w.r.t x_1 and x_2.

6.1.2 Delta method

- Suppose it is found that

$$\hat{Y} = \hat{\beta}_o + \hat{\beta}_1 X + \hat{\beta}_2 X^2$$

- How do you assign X so that Y is max/min?

$$\hat{Y} = \hat{\beta}_o - \frac{\hat{\beta}_1^{\ 2}}{4\hat{\beta}_2} + \hat{\beta}_2 \left(X + \frac{\hat{\beta}_1}{2\hat{\beta}_2} \right)^2$$

  - $\hat{X} = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$

- Can we construct confidence interval for $\hat{X}$?

6.1.2 Delta method

- Delta Method = Taylor Series + Normal approx.
- Set-up
  - Known: $\hat{\theta} \sim N(\theta, \sigma^2 D)$
  - Want : The distribution of $g(\hat{\theta})$.
- Delta Methods:
  - $g(\hat{\theta}) = g(\theta) + g'(\theta)^T(\hat{\theta} - \theta) + error$
  - $\mathrm{E}\left(g(\hat{\theta})\right) \approx g(\theta)$
  - $\mathrm{Var}\left(g(\hat{\theta})\right) \approx g'(\theta)^T \mathrm{Var}(\hat{\theta} - \theta) g'(\theta) = \sigma^2 g'(\theta)^T D g'(\theta)$

$$\boxed{g(\hat{\theta}) \sim N\left(g(\theta), \sigma^2 g'(\theta)^T D g'(\theta)\right)}$$

6.1.2 Delta method

- Delta Method

$$\hat{\theta} \sim N(\theta, \sigma^2 D)$$
$$g(\hat{\theta}) \sim N\left(g(\theta), \sigma^2 g'(\theta)^T D g'(\theta)\right)$$

- Example
  - $g(\hat{\beta}) = \hat{X} = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$, $\hat{\beta}=(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) \sim N(\beta, \sigma^2 (X'X)^{-1})$

- Delta Methods:
  - $g'(\beta) = (0, -\frac{1}{2\beta_2}, \frac{\beta_1}{2\beta_2^2})$
  - $\widehat{Var}\left(g(\hat{\beta})\right) \triangleq \hat{\sigma}^2 \left(0, -\frac{1}{2\hat{\beta}_2}, \frac{\hat{\beta}_1}{2\hat{\beta}_2^2}\right) (X'X)^{-1} \left(0, -\frac{1}{2\hat{\beta}_2}, \frac{\hat{\beta}_1}{2\hat{\beta}_2^2}\right)^T$

95% $C.I\ for g(\beta) = g(\hat{\beta}) \pm 1.96\sqrt{\widehat{Var}\left(g(\hat{\beta})\right)}$

6.1.2 Delta Method

- Suppose that $(X_1 = Time, X_2 = Temp)$

$$E(Y|X_1, X_2 = 350) = -197 + 11.3X_1 - 0.157{X_1}^2$$

$$\widehat{Var}(\hat{\beta}) = \begin{pmatrix} 3 & 1.2 & 8 \\ 1.2 & 10 & 3 \\ 8 & 3 & 3.2 \end{pmatrix} \text{x}10^{-6}$$

  - How long should we bake?

- Answer: $\hat{Y} = \hat{\beta}_o + \hat{\beta}_1 X + \hat{\beta}_2 X^2 = \hat{\beta}_o - \frac{{\hat{\beta}_1}^2}{4\hat{\beta}_2} + \hat{\beta}_2 \left(X + \frac{\hat{\beta}_1}{2\hat{\beta}_2} \right)^2$

$$g(\hat{\beta}) = - \frac{\hat{\beta}_1}{2\hat{\beta}_2} = 36 \qquad g'(\beta) = (0, - \frac{1}{2\beta_2}, \frac{\beta_1}{2\beta_2^2})$$

$$\widehat{Var}\left(g(\hat{\beta})\right) = \left(0, -\frac{1}{2\hat{\beta}_2}, \frac{\hat{\beta}_1}{2\hat{\beta}_2^2}\right) \widehat{\text{Var}}(\hat{\beta}) \left(0, -\frac{1}{2\hat{\beta}_2}, \frac{\hat{\beta}_1}{2\hat{\beta}_2^2}\right)^T = 0.17$$

95% C.I. for best baking time = $36 \pm 1.96\sqrt{0.17}$ = (35.2, 36.8)

6.2 Factors

- What is the difference between the following two groups?
  - Height, weight, speed
  - Gender, blood type, eye color.

6.2 Factors

- Two types of variables
  - Quantitative variables
    - Measured in a numeric scale.
    - Has unit
    - e.g. Height, weight, speed, population
  - Qualitative/Categorical variables
    - Divided into levels/categories
    - No unit
    - e.g. Gender, blood type, eye color.

6.2 Factors

- Simple case: Factor with 2 levels
  - e.g. male/female, pass/fail
  - Use dummy variable (also called indicator)

$$U_i = \begin{cases} 0 -- & \text{the i-th observation is F} \\ 1 -- & \text{the i-th observation is M} \end{cases}$$

  - e.g. Height$_i$ = $\beta_0+\beta_1 U_i+e_i$
    - β_1 = the additional effect for one level over another.
      = the mean height difference between male and female

6.2 Factors

- Not so simple case: Factor with >2 levels
  - e.g. blood type, eye color, college
  - e.g. GPA = $\beta_0+\beta_1 C+e$

$$C = College = \begin{cases} 0 -- & \text{CC} \\ 1 -- & \text{UC} \\ 2 -- & \text{NA} \\ 3 -- & \text{NC (New Colleges)} \end{cases}$$

- Is it reasonable?

6.2 Factors

- Factor with d levels
  - The factor rule:
    - A factor with d levels can be represented by (d-1) dummy variables and the intercept.
  - How to use dummy variable?
  - e.g. GPA = $\beta_0+\beta_1 C_{UC}+\beta_2 C_{NA}+\beta_3 C_{NC}+e$

$$C_{UC}=\begin{cases}0\text{-- not UC}\\1\text{-- UC}\end{cases},\quad C_{NA}=\begin{cases}0\text{-- not NA}\\1\text{-- NA}\end{cases},\quad C_{NC}=\begin{cases}0\text{-- not NC}\\1\text{-- NC}\end{cases}$$

    - Essentially, we model each group with a separate mean.
    - β_0=E(GPA|CC), $\beta_0+\beta_1$=E(GPA|UC), $\beta_0+\beta_2$=E(GPA|NA) , $\beta_0+\beta_3$=E(GPA|NC)

6.2 Factors

- Factor with d levels
  - e.g. GPA = $\beta_0+\beta_1 C_{UC}+\beta_2 C_{NA}+\beta_3 C_{NC}+e$

$$C_{UC}=\begin{cases}0\text{-- not UC}\\1\text{-- UC}\end{cases},\quad C_{NA}=\begin{cases}0\text{-- not NA}\\1\text{-- NA}\end{cases},\quad C_{NC}=\begin{cases}0\text{-- not NC}\\1\text{-- NC}\end{cases}$$

  - Implementation

Student	GPA	College	C_{UC}	C_{NA}	C_{NC}
Bun	3.6	UC	1	0	0
Wing	3.9	WS	0	0	1
KiKi	3.5	NA	0	1	0
Cherry	3.8	UC	1	0	0
Jeremy	3.6	CC	0	0	0
…	…	…	…	…	…

  - R-Implementation

```
College=c("uc",“nc",“na",“uc","cc")
C.uc=as.numeric(College=="uc"); C.na=as.numeric(College=="na"); C.nc=as.numeric(College==“nc")
```

6.2 Factors

- Example (Sleep data, sleep1.txt)
  - Y = total hours of sleep
  - X = D = danger of species = 1 to 5.

Constant variance assumption ok

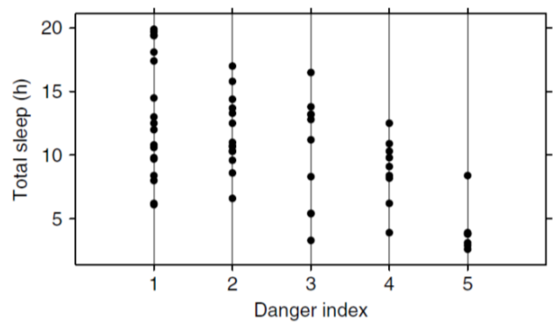


FIG. 6.5 Total sleep versus danger index for the sleep data.

- Model:
  - $E(TS|D) = \eta_0 + \eta_2 U_2 + \eta_3 U_3 + \eta_4 U_4 + \eta_5 U_5$
- Result:

	Estimate	Std. Error	t-value	Pr(>|t|)
(b) Mean function (6.16)				
Intercept	13.0833	0.8881	14.73	0.0000
U_2	−1.3333	1.3427	−0.99	0.3252
U_3	−2.7733	1.4860	−1.87	0.0675
U_4	−4.2722	1.5382	−2.78	0.0076
U_5	−9.0119	1.6783	−5.37	0.0000

	Df	Sum Sq	Mean Sq	F-value	Pr(>F)
D	4	457.26	114.31	8.05	0.0000
Residuals	53	752.41	14.20		

6.2 Factors

- Remark 1
  - We may use the model

    Convenient for interpretable parameters

    - $E(TS|D) = \beta_1 U_1 + \beta_2 U_2 + \beta_3 U_3 + \beta_4 U_4 + \beta_5 U_5$

    Instead of

    Convenient for comparing mean functions

    - $E(TS|D) = \eta_0 + \eta_2 U_2 + \eta_3 U_3 + \eta_4 U_4 + \eta_5 U_5$

    The major difference is in the interpretation of estimates:

    β_k= the mean of group k

    η_k= the mean of group k over group 1
- Remark 2
  - This is an example of One way ANOVA
    - Fit a separate mean for each level of a factor.
      - Null model -- H_o: All groups have the same mean, i.e. η_s=0, for s>0
      - Alternative -- H_A: Not all groups have the same mean, i.e. some $\eta_s \neq 0$

6.2 General Cases – add more predictors

- Example (Sleep data, sleep1.txt)
  - Y = total hours of sleep
  - X_1 = D = danger of species = 1 to 5.
  - X_2 = Body weight
- Model 1 (Full model)

$$E(TS|\log(BodyWt) = x, D = j) = \eta_0 + \eta_1 x + \sum_{i=2}^{d} (\eta_{0j} U_j + \eta_{1j} U_j x)$$

- Model 2 (Parallel regression)

$$E(TS|\log(BodyWt) = x, D = j) = \eta_0 + \eta_1 x + \sum_{j=2}^{d} (\eta_{0j} U_j)$$

- Model 3 (Common intercept)

$$E(TS|\log(BodyWt) = x, D = j) = \eta_0 + \eta_1 x + \sum_{j=2}^{d} (\eta_{1j} U_j x)$$

- Model 4 (simple regression, ignore effect of D)

$$E(TS|\log(BodyWt) = x, D = j) = \eta_0 + \eta_1 x$$

6.2 General Cases – add more predictors

- Implementation
- Model 1 (Full model)

$$\mathrm{E}(TS|\log(BodyWt) = x, D = j) = \eta_0 + \eta_1 x + \sum_{j=2}^{d} (\eta_{0j} U_j + \eta_{1j} U_j x)$$

Animal	TS	D	U_2	U_3	U_4	U_5	X (BodyWt)	U_2X	U_3X	U_4X	U_5X
1	4.1	2	1	0	0	0	51	51	0	0	0
2	5.3	1	0	0	0	0	62	0	0	0	0
3	3.0	3	0	1	0	0	23	0	23	0	0
4	8.7	5	0	0	0	1	70	0	0	0	70
…	…	…	…							…	…
…	…	…	…							…	…

6.2 General Cases – add more predictors

- Model 1 (Full model)

$$E(TS|\log(BodyWt) = x, D = j) = \eta_0 + \eta_1 x + \sum_{j=2}^{d} (\eta_{0j} U_j + \eta_{1j} U_j x)$$

- Model 2 (Parallel regression)

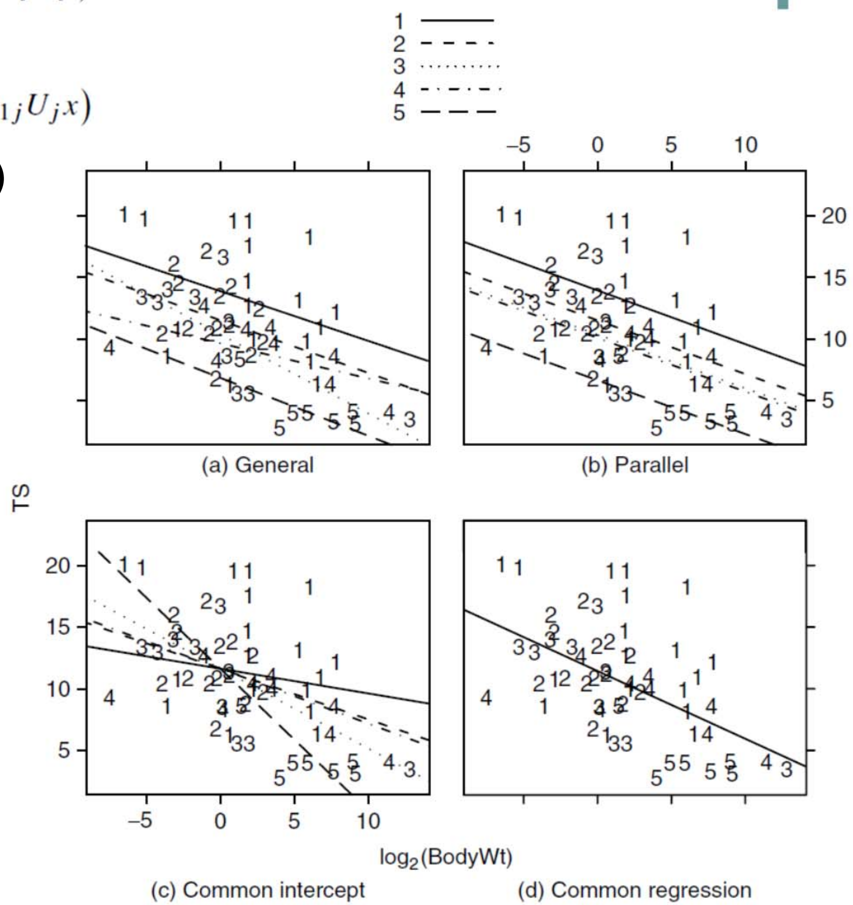
$$E(TS|\log(BodyWt) = x, D = j) = \eta_0 + \eta_1 x + \sum_{j=2}^{d} (\eta_{0j} U_j)$$

- Model 3 (Common intercept)

$$E(TS|\log(BodyWt) = x, D = j) = \eta_0 + \eta_1 x + \sum_{j=2}^{d} (\eta_{1j} U_j x)$$

- Model 4 (simple regression, ignore effect of D)

$$E(TS|\log(BodyWt) = x, D = j) = \eta_0 + \eta_1 x$$



- Model Selection using F- test:

TABLE 6.2 Residual Sum of Squares and df for the Four Mean Functions for the Sleep Data

	df	RSS	F	P(>F)
Model 1, most general	48	565.46		
Model 2, parallel	52	581.22	0.33	0.853
Model 3, common intercept	52	709.49	3.06	0.025
Model 4, all the same	56	866.23	3.19	0.006

$$\boxed{F_\ell = \frac{(RSS_\ell - RSS_1)(\mathrm{df}_\ell - \mathrm{df}_1)}{RSS_1/\mathrm{df}_1}}$$