

## 4. Modelos de Regressão Múltipla

### 4.1 Regressão com replicação

#### 4.1.1 O Modelo

$$Y_i \overset{ind}{\sim} N(\beta_1 + \beta_2 x_i, \sigma^2), \quad i = 1, \dots, n \quad (1)$$

Princípio da Parcimônia: Se dois modelos explicam igualmente bem um determinado fenômeno, escolha aquele que envolve um número menor de parâmetros.

Por exemplo, se em (1) a hipótese  $\beta_2 = 0$  é aceita, então, pelo PP

$$Y_i \overset{ind}{\sim} N(\beta_1, \sigma^2), \quad i = 1, \dots, n$$

Para o modelo (1) podemos frequentemente obter replicações do experimento para cada valor de  $x$  tal que

$$Y_{ij} \stackrel{ind}{\sim} N(\beta_1 + \beta_2 x_i, \sigma^2), \quad i = 1, \dots, k; j = 1, \dots, n_i \quad (2)$$

O modelo (2) é um submodelo do modelo a um fator estudado no Cap. 3 dado por

$$Y_{ij} \stackrel{ind}{\sim} N(\delta_i, \sigma^2), \quad i = 1, \dots, k; j = 1, \dots, n_i \quad (3)$$

Em particular, para cada um dos  $k$  grupos indexados por  $i$ , temos  $n_i$  replicações iid's, resultando em um total de  $n = \sum_i n_i$  observações.

Vantagens de (2) comparado com (1):

(i) podemos obter o estimador de  $\sigma^2$  baseado em (3) que incorpora a replicação e não envolve suposições sobre a forma da relação entre  $Y$  e  $x$ .

(ii) replicação permite uma estimação mais precisa dos  $\delta_i$ 's em (3) dando maior poder ao teste de (2) sobre (3).

O Modelo (2) é chamado modelo de regressão linear com replicação (MRL c/rep.)

Ambos (2) e (3) são ML's e como (2) é obtido de (3) supondo  $\delta_i = \beta_1 + \beta_2 x_i$ ,  $i = 1, \dots, n$ , segue que (2) é um submodelo de (3).

## ESTRATÉGIA:

(i) Verificar (3) usando análise de resíduos e outras técnicas já discutidas. Realizar o teste de Bartlett para verificar a igualdade das variâncias.

(ii) Aceito o modelo (3), podemos testar o modelo (2) sob (3), usando um teste  $F$ .

Adotando a notação do Cap. 2, denote o modelo (3) por  $\mu \in L_1$  e o modelo (2) por  $\mu \in L_2 \subset L_1$ .

Estimação dos parâmetros do modelo  $L_1$ :

$$\hat{\delta}_i = \bar{y}_{i+}, \quad i = 1, \dots, k \text{ e}$$

$$\tilde{\sigma}_1^2 = \frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i+})^2.$$

Estimação dos parâmetros em  $L_2$ .

O modelo (2) é um caso especial do modelo (1) obtido supondo-se que certos valores de  $x_i$  são iguais. Logo, usando os resultados do Cap. 1 e adotando a notação lá utilizada, podemos obter os estimadores de  $\beta_1$ ,  $\beta_2$  e  $\sigma^2$  em  $L_2$ .

Seja  $\bar{x}_{++} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij} = \frac{1}{n} \sum_{i=1}^k n_i x_i$  e  $t_i = x_i - \bar{x}_{++}$ ,  $i = 1, \dots, k$ . Defina  $t \in \mathbb{R}^n$  como  $t = (\underbrace{t_1, \dots, t_1}_{n_1}, \underbrace{t_2, \dots, t_2}_{n_2}, \dots, \underbrace{t_k, \dots, t_k}_{n_k})^T$ .

Tomando  $\mu_{ij} = E[Y_{ij}]$ , escrevemos, sob  $L_2$ ,  $\mu_{ij} = \beta_1 + \beta_2 x_i = \alpha + \beta_2 t_i$ , onde  $\alpha = \beta_1 + \beta_2 \bar{x}_{++}$ .

Usando o fato de que  $\underline{1} \perp t$  obtemos

$\hat{\alpha} = \bar{y}_{++} = \frac{1}{n} \sum_i \sum_j y_{ij}$  e  $\hat{\beta}_2 = \frac{S_{ty}}{S_t}$ , onde  $S_{ty} = \sum_i \sum_j t_i y_{ij}$  e  $S_t = \sum_i n_i t_i^2$ .

Assim,  $\hat{\beta}_1 = \bar{y}_{++} - \bar{x}_{++}\hat{\beta}_2$  e o estimador de  $\sigma^2$  sob  $L_2$  é

$$\tilde{\sigma}_2^2 = \frac{1}{n-2} \sum_i \sum_j (y_{ij} - \hat{\alpha} - \hat{\beta}_2 t_i)^2$$

Derivação do teste  $F$  de  $L_2$  sob  $L_1$ :

$$\|p_1(y) - p_2(y)\|^2 = \sum_i \sum_j (\bar{y}_{i+} - \hat{\alpha} - \hat{\beta}_2 t_i)^2 = \sum_i n_i ((\bar{y}_{i+} - \hat{\alpha} - \hat{\beta}_2 t_i)^2).$$

Observe que esta é uma soma dos quadrados ponderados das diferenças entre os estimadores de  $\delta_i$  sob (3) e sob (2).

O teste  $F$  para a regressão linear é então

$$F(y) = \frac{\sum_i n_i ((\bar{y}_{i+} - \hat{\alpha} - \hat{\beta}_2 t_i)^2 / (k-2))}{\tilde{\sigma}_1^2}$$

Sob  $L_2$ ,  $F(Y) \sim F_{(k-2, n-k)}$ .

Se aceitamos  $L_2$ , podemos continuar a análise testando  $\beta_2 = 0$  usando um teste  $t$ .

Suponha então que  $L_2$  foi aceito. Seja  $H_0 : \beta_2 = 0$ . Vimos que  $\hat{\beta}_2 = \frac{S_{ty}}{S_t}$ . Seja  $t(y) = \frac{\hat{\beta}_2}{\tilde{\sigma}_2/\sqrt{S_t}}$ . Sob  $H_0$ ,  $t(Y) \sim t_{(n-2)}$ .

## 4.2 Comparação de linhas de regressão

### 4.2.1 Teoria

Extensão do MRL simples: comparação de várias linhas de regressão (Análise de covariância)

Suponha um dado fator  $I$  e um MRL simples para cada nível de  $I$

$$Y_{ij} \stackrel{ind}{\sim} B(\beta_{1i} + \beta_{2i}x_{ij}, \sigma^2), \quad i = 1, \dots, k; j = 1, \dots, n_i \quad (4)$$

onde  $n = \sum_i n_i$ . Os  $2k$  coeficientes das  $k$  linhas de regressão podem ser diferentes e os conjuntos dos valores de  $x$  podem variar de linha para linha.

Observe também que (4) pode ser estendido de forma a incluir replicação. Porém aqui, consideraremos apenas o caso sem replicação.

Seja  $\mu = (\mu_{11}, \dots, \mu_{1n_1}, \dots, \mu_{k1}, \dots, \mu_{kn_k})^T$ . Então, sob (4) temos que  $\mu = \sum_i (e_i \beta_{1i} + u_i \beta_{2i})$ , onde  $e_i = (0, \dots, 0, \underbrace{1, \dots, 1}_{\text{grupo } i}, 0, \dots, 0)^T$  e

$$u_i = (0, \dots, 0, \underbrace{x_{i1}, \dots, x_{in_i}}_{\text{grupo } i}, 0, \dots, 0)^T$$

Como os  $\beta$ 's variam livremente no modelo, o modelo é linear e é dado por

$$L_1 = \text{span}\{e_1, \dots, e_k, u_1, \dots, u_k\}.$$



Por exemplo, se  $k = 2$  temos

$$\mu = \beta_{11} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \beta_{12} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_{21} \begin{pmatrix} x_{11} \\ \vdots \\ x_{1n_1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \beta_{22} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ x_{21} \\ \vdots \\ x_{2n_2} \end{pmatrix}$$

Para obter uma base ortogonal para  $L_1$  observe primeiro que  $e_i \cdot e_j = 0$ ,  $u_i \cdot u_j = 0$  e  $e_i \cdot u_j = 0$ ,  $\forall i \neq j$ .

Assim, basta ortogonalizar o par  $(e_i, u_i)$  para cada  $i$ .

Defina  $t_i = u_i - \bar{x}_{i+} e_i$  onde  $\bar{x}_{i+} = \frac{1}{n_i} \sum_j x_{ij}$ .

Então  $t_i$  é função de  $e_i$  e  $u_i$  e  $t_i \cdot e_i = e_i \cdot u_i - e_i \cdot e_i \bar{x}_{i+} = x_{i+} - n_i \bar{x}_{i+} = 0$ , onde  $x_{i+} = \sum_j x_{ij}$ .

Assim,  $e_1, \dots, e_k, t_1, \dots, t_k$  é uma base ortogonal para  $L_1$ . Na verdade, estamos realizando uma regressão linear separadamente para cada um dos  $k$  grupos.

Trabalhando com o novo sistema de coordenadas temos

$\mu_{ij} = \sum_i (e_i \alpha_i + t_i \beta_{2i})$  onde  $\alpha_i = \beta_{1i} + \beta_{2i} \bar{x}_{i+}$  e, pela ortogonalidade da base, temos

$$\hat{\alpha}_i = \frac{y \cdot e_i}{||e_i||^2} = \frac{1}{n_i} \sum_j y_{ij} = \bar{y}_{i+}$$

$$\hat{\beta}_{2i} = \frac{y \cdot t_i}{||t_i||^2} = \frac{1}{S_{t_i}} \sum_j y_{ij} t_{ij} = \frac{S_{ty_i}}{S_{t_i}}$$

onde  $S_{yt_i} = \sum_j y_{ij} t_{ij}$  e  $S_{t_i} = \sum_j t_{ij}^2$

$$\hat{\beta}_{1i} = \hat{\alpha}_i - \hat{\beta}_{2i}\bar{x}_{i+}.$$

Um estimador de  $\sigma^2$  baseado nas observações do grupo  $i$  é

$$\tilde{\sigma}_{(i)}^2 = \frac{1}{n_i - 2} \sum_j (y_{ij} - \hat{\beta}_{1i} - \hat{\beta}_{2i}x_{ij})^2.$$

$\tilde{\sigma}_{(1)}^2, \dots, \tilde{\sigma}_{(k)}^2$  são independentes.

Podemos aplicar o teste de Bartlett para comparar as variâncias em cada grupo.

O estimador de  $\sigma^2$  sob  $L_1$  é obtido combinando-se os estimadores  $\tilde{\sigma}_{(i)}^2$ ,  $i=1, \dots, k$  resultando em

$$\tilde{\sigma}_1^2 = \frac{1}{n-2k} \sum_i \sum_j (y_{ij} - \hat{\beta}_{1i} - \hat{\beta}_{2i}x_{ij})^2 = \frac{1}{f_+} \sum_i f_i \tilde{\sigma}_{(i)}^2$$

$f_i = n_i - 2$ ,  $f_+ = \sum_i f_i$  e  $\tilde{\sigma}_1^2$  com  $n - 2k$  graus de liberdade.

Verificação do modelo:

diagrama de dispersão de  $y$  versus  $x$  para cada grupo separadamente (;inearidade).

analisar os resíduos do modelo definidos por  $r_{ij} = y_{ij} - \hat{\beta}_{1i} - \hat{\beta}_{2i}x_{ij}$ .

gráfico de  $r_{ij}$  versus  $i$  ou valores ajustados (homogeneidade das variâncias)

normal-plot dos resíduos.

Usando a notação de Wilkinson e Rogers, definimos a interação entre um fator  $I$  e um vetor  $x$  pelo conjunto

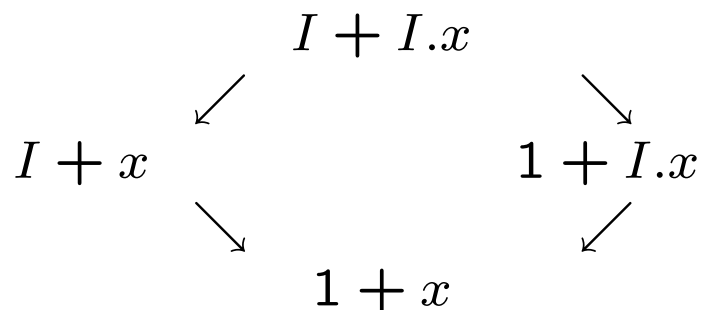
$$I.x = \{e_1x, \dots, e_kx\},$$

onde  $e_1, \dots, e_k$  são os vetores dummy correspondentes à  $I$  e  $e_ix$  representa o produto das coordenadas equivalentes de  $e_i$  e  $x$ .

Assim, a fórmula do modelo tratado aqui é

$$I + I.x$$

Um diagrama de modelo aqui é



Em geral, é melhor testar primeiro se a inclinação é comum pois esta hipótese não depende da escolha da origem no eixo  $x$ . Testar primeiro o intercepto comum pode não fazer sentido a não ser que a origem do eixo  $x$  tenha um significado especial no contexto do experimento que gerou os dados.

$$I + I.x \longrightarrow I + x \longrightarrow 1 + x$$

Primeiro testamos  $H_2 = \beta_{21} = \beta_{22} = \dots = \beta_{2k}$  sob  $\mu \in L_1$ . Se  $H_2$  é aceita podemos testar  $H_3 : \beta_{11} = \beta_{12} = \dots = \beta_{1k}$  sob  $H_2$ .

Sob  $H_3$  o modelo é  $Y_{ij} \stackrel{ind}{\sim} N(\beta_1 + \beta_2 x_{ij}, \sigma^2)$  que é o MRL simples.

Ambas as hipóteses  $H_2$  e  $H_3$  são ML's.

Denote por  $L_2$  e  $L_3$  os espaços lineares correspondentes. Como já vimos é útil trabalharmos com uma base ortogonal de  $L_1$ , por exemplo, a base  $\{e_1, \dots, e_k, t_1, \dots, t_k\}$ .

Assim, uma base ortogonal para  $L_2$  é  $\{e_1, \dots, e_k, t_+\}$  onde  $t_+ = \sum_i t_i$ .

Uma base ortogonal para  $L_3$  é  $\{1, v\}$  onde  $v = u_1 + \dots + u_k - \bar{x}_+ + 1$ .

Os estimadores sob  $L_1$  já foram obtidos.

Sob  $L_2$  os estimadores são  $\hat{\alpha}_i = \bar{y}_{i+}$ ,  $i = 1, \dots, k$   
e  $\hat{\beta}_2^{(2)} = \frac{y \cdot t_+}{\|t_+\|^2} = \frac{S_{ty+}}{S_{t+}}$  onde

$$S_{ty+} = \sum_i \sum_j y_{ij} t_{ij} \text{ e } S_{t+} = \sum_i \sum_j t_{ij}^2.$$

O estimador para  $\sigma^2$  sob  $H_2$  é

$$\tilde{\sigma}_2^2 = \frac{1}{n-k-1} \sum_i \sum_j (y_{ij} - \hat{\alpha}_i - \hat{\beta}_2^{(2)} t_{ij})^2.$$

Sob  $L_3$  os estimadores são  $\hat{\alpha} = \bar{y}_{++}$ ,  $i = 1, \dots, k$   
e  $\hat{\beta}_2^{(3)} = \frac{S_{vy}}{S_v}$  onde

$$S_{vy} = \sum_i \sum_j y_{ij} v_{ij} \text{ e } S_v = \sum_i \sum_j v_{ij}^2.$$

O estimador para  $\sigma^2$  sob  $L_3$  é

$$\tilde{\sigma}_3^2 = \frac{1}{n-2} \sum_i \sum_j (y_{ij} - \hat{\alpha} - \hat{\beta}_2^{(3)} v_{ij})^2.$$

O teste  $F$  de  $H_2$  sob  $H_1$  é

$$F(y) = \frac{\sum_i \sum_j t_{ij}^2 (\hat{\beta}_{2i} - \hat{\beta}_2^{(2)})^2 / (k-1)}{\tilde{\sigma}_1^2}$$

Sob  $L_2$ ,  $F(Y) \sim F_{(k-1, n-2k)}$ .

O teste  $F$  de  $H_3$  sob  $H_2$  é

$$F(y) = \frac{\sum_i \sum_j (\hat{\alpha}_i + \hat{\beta}_2^{(2)} t_{ij} - \hat{\alpha} - \hat{\beta}^{(3)} t_{ij})^2 / (k-1)}{\tilde{\sigma}_2^2}$$

Sob  $L_3$  dado  $L_2$ ,  $F(Y) \sim F_{(k-1, n-k-1)}$ .

Os testes para a sequência  $I + I.x \rightarrow 1 + I.x \rightarrow 1 + x$  são igualmente fáceis de serem obtidos.

(EXEMPLO: UFFI data)



## 4.3 Regressão Múltipla

### 4.3.1 Modelo

$$Y_i \stackrel{ind}{\sim} N(\mu_i, \sigma^2), \quad i = 1, \dots, n$$

onde  $\mu_i = \beta_1 x_{i1} + \dots + \beta_k x_{ik}$ .

Os  $x_{ij}$ 's são conhecidos,  $\beta_1, \dots, \beta_k$  são parâmetros desconhecidos.

Na forma matricial:  $\mu = X\beta$ , onde  $X$  é uma matriz  $n$  por  $k$ ,  $\beta^T = (\beta_1, \dots, \beta_k)$  e  $\mu^T = (\mu_1, \dots, \mu_k)$ .

Frequentemente, mas não necessariamente, a primeira coluna de  $X$  é uma coluna de 1's.

Este modelo é uma forma mais geral do ML e inclui como caso especial todos os modelos já apresentados até aqui.

Geralmente,  $x_1, \dots, x_k$  são variáveis contínuas.

Explicar  $Y$  por uma função linear dos  $x_j$ 's.

Supomos que as colunas de  $X$  são li's fazendo  $\beta_1, \dots, \beta_k$  a única representação de  $\mu$  para  $\mu \in L_1$  onde  $L_1 = \text{span}\{x_1, \dots, x_k\}$ . Então o estimador de  $\beta$  é  $\hat{\beta} = (X^T X)^{-1} X^T y$  e, pelo teorema de Pitágoras,

$$\tilde{\sigma}_1^2 = \frac{1}{n-k} \{ \|y\|^2 - \|X\hat{\beta}\|^2 \} = \frac{1}{n-k} \{ \|y\|^2 - \hat{\beta}^T X^T X \hat{\beta} \}$$

A distribuição de  $\hat{\beta}$  sob  $L_1$  é  $\hat{\beta} \sim N_k(\beta, \sigma^2 C)$ , onde  $C = (X^T X)^{-1} = (c_{ij})$ . Lembre que estes resultados foram obtidos no Cap. 2. Também vimos que o teste  $t$  para a hipótese  $\beta_j = 0$  é dado por

$$t(y) = \frac{\hat{\beta}_j}{\tilde{\sigma}_1 c_{jj}^{1/2}}$$

tal que sob  $\beta_j = 0$ ,  $t(Y) \sim t_{(n-k)}$ .

Em geral, o interesse está em mais de um  $\beta_j$  ao mesmo tempo.

Considere então a hipótese  $H_2 : \beta_{l+1} = \dots = \beta_k = 0$  que é uma hipótese linear dada por  $\mu \in L_2 = \text{span}\{x_1, \dots, x_l\}$ .

Particione a matriz  $X$  como  $X = [X_1 | X_2]$  tal que  $X_1 = [x_1 \dots x_l]$  e  $X_2 = [x_{l+1} \dots x_k]$ .

Sob  $L_2$ , e representando  $\mu$  como  $X_1\psi$ ,  $\psi \in \mathbb{R}^l$  temos  $\hat{\psi} = (X_1^T x_1)^{-1} X_1^T y$ .

Escrevendo  $X\beta = X_1\beta^{(1)} + X_2\beta^{(2)}$  temos

$$\|X\hat{\beta} - X_1\hat{\psi}\|^2 = \|X_1(\hat{\beta}^{(1)} - \hat{\psi}) + X_2\hat{\beta}^{(2)}\|^2.$$

Assim, o teste  $F$  de  $L_2$  sob  $L_1$  é

$$F(y) = \frac{\|X_1(\hat{\beta}^{(1)} - \hat{\psi}) + X_2\hat{\beta}^{(2)}\|^2 / (k-l)}{\tilde{\sigma}_1^2}$$

Sob  $L_2$ ,  $F(Y) \sim F_{(k-l, n-k)}$ .

No caso especial onde  $X_1^T X_2$  é uma matriz nula, isto é os espaços  $L_2$  e  $L_{2'} = \text{span}\{x_{l+1}, \dots, x_k\}$  são ortogonais, temos que

$X_1\hat{\beta}^{(1)}$  é a projeção ortogonal sobre  $L_2$  dando  $\hat{\beta}^{(1)} = \hat{\psi}$ . Neste caso, a estatística  $F$  simplifica para

$$F(y) = \frac{\|X_2\hat{\beta}^{(2)}\|^2 / (k-l)}{\tilde{\sigma}_1^2} = \frac{\hat{\beta}^{(2)T} X_2^T X_2 \hat{\beta}^{(2)} / (k-l)}{\tilde{\sigma}_1^2}$$

Obs.: Qualquer teste  $F$  pode ser colocado na forma acima para alguma escolha adequada da base para  $L_1$ . Porém, o ponto de distinção aqui é que no caso típico de regressão múltipla, as variáveis  $x$  são dadas e a base, assim, não é aberta a escolhas.

(EXEMPLO: Tree data)

## 4.4 Regressão Polinomial

### 4.4.1 Polinômios em uma variável

⇒ Relações complexas não-lineares entre uma variável dependente  $Y$  e uma variável independente  $x$  podem ser frequentemente descritas por uma relação polinomial.

Considere o modelo

$$Y_i = \beta_1 + \beta_2 x_1 + \dots + \beta_{k+1} x_k + \epsilon_i$$

onde  $\epsilon_i \stackrel{ind}{\sim} N(0, \sigma^2)$ .

Observe que enquanto o modelo acima descreve uma relação não-linear entre  $Y$  e  $x$ , este é de fato um modelo linear com variáveis explicativas dadas por  $x_{ij} = x_i^j$ .

É claro que devemos ter  $k + 1 < n$  de forma a ter um modelo significativo e de fato, trabalha-se com valores de  $k$  pequenos tais como 2 ou 3. Note que o caso  $k = n - 1$  é o modelo saturado.

O MRP apresenta certos problemas em sua análise e interpretação pois é muito usado para propósitos descritivos. Assim, se  $f(x)$  é uma função  $k$  vezes diferenciável podemos aproximar  $f$  por sua série de Taylor de ordem  $k$  em torno de zero, que escrevemos, sugestivamente, na forma,

$$f(x) \simeq \beta_1 + \beta_2 x + \dots \beta_{k+1} x^k$$

Deste ponto de vista, o modelo polinomial pode ser usado para aproximar uma relação não linear arbitrária, porém “regular”, entre  $E[Y]$  e  $x$ .

Esta técnica pode ser muito útil. Porém, um dos perigos da regressão polinomial está na extrapolação. Mesmo que a aproximação por um polinômio de ordem  $k$  seja boa em um dado intervalo de  $x$ , ela pode não fazer sentido fora deste intervalo.

Conclusões baseadas sobre a análise do modelo polinomial com uma variável não devem ser estendidas para valores de  $x$  fora do intervalo disponível.

Um problema potencial com a análise do MRP é que pode-se encontrar grandes correlações entre os vetores  $x^j$ .

No caso onde a relação entre  $E[Y]$  e  $x$  é monótona, um polinômio de regressão alternativo é uma regressão com uma variável explicativa não linear tal como  $\log(x)$ ,  $e^x$  ou  $x^\alpha$ .

Isto pode levar a um modelo mais parcimonioso pois uma única variável explicativa é usada para construir um modelo para uma relação não-linear.



Assim, um modelo tal como  $\beta_1 + \beta_2 \log x$  pode ser substituído por uma regressão polinomial em  $x$ , mas será necessário pelo menos um termo quadrático  $\beta_1 + \beta_2 x + \beta_3 x^2$  para obter um ajuste comparável.

A escolha prática entre tais alternativas pode ser feita via análise dos resíduos.

#### 4.4.2 Polinômios Ortogonais

Usar polinômios ortogonais significa ortogonalizar a base para o espaço linear dos polinômios, evitando assim o problema de dependência linear aproximada da base mencionada anteriormente.

Lembre que a forma ortogonal do MRL é:

$$Y_i = \alpha + \beta_2(x_i - \bar{x}_+) + \epsilon_i$$

que leva a estimadores independentes  $\hat{\alpha} = \bar{Y}_+$  e  $\hat{\beta}_2 = \frac{S_{ty}}{S_t}$ .

Para um ML com uma base ortogonal  $\{x_1, \dots, x_k\}$  temos estimadores independentes  $\hat{\beta}_j = \frac{x_j \cdot y}{\|x_j\|^2}$  com variâncias  $\frac{\sigma^2}{\|x_j\|^2}$ .

Além disso, podemos escrever o estimador de  $\sigma^2$  como

$$\tilde{\sigma}^2 = \frac{1}{n-k} \{S_y - \sum_j \hat{\beta}_j^2 \|x_j\|^2\} = \frac{1}{n-k} \{S_y - \sum_j \hat{\beta}_j^2 \sum_i x_{ij}^2\}$$

Vejamos primeiro um caso simples de regressão quadrática onde a variável explicativa assume os valores  $0, \pm 1, \pm 2, \dots, \pm r$  simetricamente distribuídos em torno de zero.

$$Y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \epsilon_i$$

Podemos escrever este modelo como  $\mu \in L = \text{span}\{1, x, x^2\}$ .

Para esta escolha particular dos valores de  $x$  temos que  $1 \perp x$  e  $x \perp x^2$ . A projeção ortogonal de  $x^2$  sobre 1 é:

$$\frac{1 \cdot x^2}{\|1\|^2} = \frac{S_x}{n} 1$$

Podemos então escrever o modelo na forma

$$Y_i = \alpha_0 + \beta_2 x_i + \beta_3 \left( x_i^2 - \frac{S_x}{n} \right) \text{ onde } \alpha_0 = \beta_1 + \beta_3 \frac{S_x}{n}.$$

Usando o fato de que agora temos uma base ortogonal segue que os estimadores independentes são

$$\hat{\alpha}_0 = \bar{Y}_+, \quad \hat{\beta}_2 = \frac{S_{xy}}{S_x} \quad \text{e} \quad \hat{\beta}_3 = \frac{\sum_i Y_i x_i^2 - \bar{Y}_+ S_x}{SSS_x} \quad \text{onde} \\ SSS_x = \sum_i (x_i^2 - \frac{1}{n} S_x)^2.$$

$$\text{Além disso, } Var(\hat{\alpha}_0) = \frac{\sigma^2}{n}, \quad Var(\hat{\beta}_2) = \frac{\sigma^2}{S_x} \quad \text{e} \\ Var(\hat{\beta}_3) = \frac{\sigma^2}{SSS_x}$$

e

$$\tilde{\sigma}^2 = \frac{1}{n-3} \left\{ S_y - n\bar{Y}_+^2 - \hat{\beta}_2 S_{xy} - \hat{\beta}_3 \sum_i (x_i^2 - S_x/n) Y_i \right\}$$

No caso geral, para valores arbitrários de  $x_1, \dots, x_n$ , escrevemos o polinômio em termos de  $k + 1$  polinômios ortogonais  $P_1, \dots, P_k$  tal que

$$Y_i = \sum_{j=0}^k \alpha_j P_j(x_i) + \epsilon_i$$

$i = 1, \dots, n$  onde  $P_0(x) = 1$  e, pela ortogonalidade  $\sum_i P_r(x_i) P_s(x_i) = 0$ ,  $r \neq s$ .

Ainda devido à ortogonalidade, segue que  $\hat{\alpha}_j = \frac{\sum_i P_j(x_i) Y_i}{\|P_j(x)\|^2}$  onde  $\|P_j(x)\|^2 = \sum_i P_j^2(x_i)$ .

Em particular,  $\hat{\alpha}_0 = \bar{Y}_+$ .

Além disso,

$$\hat{\alpha}_j \sim N\left(\alpha_j, \frac{\sigma^2}{\|P_j(x)\|^2}\right)$$

$$\tilde{\sigma}^2 = \frac{1}{n-k-1} \{S_y - \sum_j \hat{\alpha}_j \sum_i P_j(x_i) Y_i\}$$

A forma real dos polinômios ortogonais depende do conjunto dos  $x_i$ 's escolhido.

Como vimos acima, para escolhas particulares dos valores  $x_i$ 's a forma dos polinômios ortogonais pode ser bem simples, mas em geral, deve ser encontrada pelo método de ortogonalização de Gram-Schmidt aplicado aos vetores  $1, x, x^2, \dots, x^k$ .

Os três primeiros polinômios ortogonais no caso geral são  $P_0(x) = 1$ ,  $P_1(x) = x - \bar{x}_+$  e

$$P_2(x) = x^2 - x \frac{S_x}{n} - \frac{\sum_i x_i^2 (x_i - \bar{x}_+)}{SS_x}, \text{ onde } SS_x = \sum_i (x_i - \bar{x}_+)^2.$$

O passo geral neste processo fornece  $P_j$  em função de  $P_0, P_1, \dots, P_{j-1}$ :

$$P_j(x) = x^j - \sum_{l=1}^{j-1} P_l(x) \frac{\sum_i x_i^j P_l(x_i)}{\sum_i P_l^2(x_i)}$$

É fácil ver que  $\alpha_k = \beta_k$  de forma que podemos derivar o teste  $t$  para a hipótese  $\beta_k = 0$ .

### 4.4.3 Duas variáveis

Consideremos a forma geral do polinômio quadrático com duas variáveis  $x_1$  e  $x_2$ .

$$q(x_1, x_2) = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_1 x_2$$

ao qual nos referimos como uma superfície de resposta. Esta é uma família de funções bastante flexível para descrever superfícies bidimensionais devido ao fato de que ela pode tomar diversas formas diferentes, dependendo dos valores dos coeficientes na forma quadrática.

se  $q$  é positiva definida, então  $q$  tem um mínimo

se  $q$  é negativa definida então  $q$  tem um máximo

se  $q$  não é definida então  $q$  tem um ponto de sela.

O modelo estatístico correspondente é

$$Y_i \stackrel{ind}{\sim} N(q(x_{i1}, x_{i2}), \sigma^2), \quad i = 1, \dots, n$$

Tal modelo pode ser útil para descrever relações entre  $Y$  e  $x_1, x_2$ ) complexas e, possivelmente não lineares.

Em particular, se  $Y$  é a eficiência de algum processo de produção, o método pode ser usado para maximizar a eficiência do processo como uma função de  $(x_1, x_2)$  localizando-se o máximo da superfície de resposta ajustada.

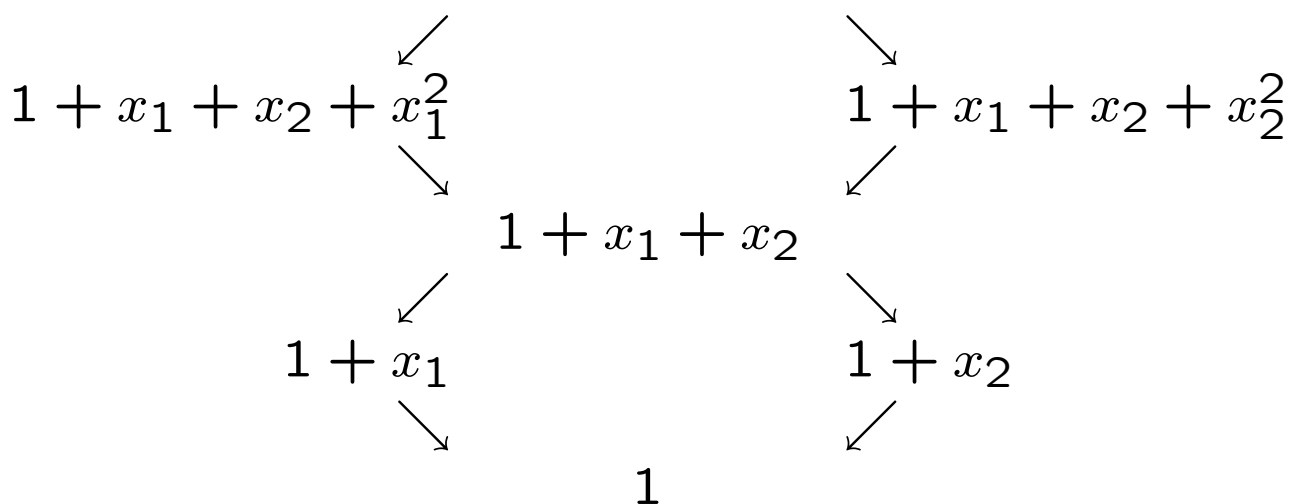


DIAGRAMA DE MODELO para a superfície de resposta ajustada para os submodelos mais importantes.

$$1 + x_1 + x_2 + x_1^2 + x_2^2 + x_1x_2$$

↓

$$1 + x_1 + x_2 + x_1^2 + x_2^2$$



## 4.5 Intervalos de Predição

### 4.5.1 Regressão Linear Simples

Consideraremos agora dois problemas similares, mas não equivalentes, em regressão linear.

⇒ Previsão de  $\mu = E[Y]$  para um novo  $x$ .

⇒ Previsão do valor de uma observação futura para um valor de  $x$ .

Suponha o modelo de regressão linear simples:

$$Y_i \stackrel{ind}{\sim} N(\beta_1 + \beta_2 x_i, \sigma^2), \quad i = 1, \dots, n$$

Seja  $x_0$  um dado valor de  $x$  não necessariamente escolhido entre os valores  $x_1, \dots, x_n$ .

Seja  $\mu_{x_0} = \beta_1 + \beta_2 x_0$  o valor de  $E[Y]$  para uma observação futura de  $Y$  com  $x = x_0$ .

Como  $\mu_{x_0}$  é uma função linear de  $\beta_1$  e  $\beta_2$ , a questão é simplesmente obter um intervalo de confiança para um dado parâmetro.

De fato, podemos reparametrizar o modelo em termos de  $\mu_{x_0}$  e qualquer outro parâmetro adequado, por exemplo,  $\beta_2$  e, então usar os métodos padrões para fazer inferência.

Porém, não há necessidade de reparametrizar o modelo em termos de  $\mu_{x_0}$  pois, na verdade, o estimador de  $\mu_{x_0}$  é

$$\hat{\mu}_{x_0} = \hat{\beta}_1 + \hat{\beta}_2 x_0 = \hat{\alpha} + \hat{\beta}_2 (x_0 - \bar{x}_+)$$

onde  $\hat{\alpha} = \bar{Y}_+$  e  $\hat{\beta}_2 = \frac{S_{ty}}{S_t}$ .

Devido à independência entre  $\hat{\alpha}$  e  $\hat{\beta}_2$  temos que

$$Var(\hat{\mu}_{x_0}) = \sigma^2 \left( \frac{1}{n} + \frac{(x_0 - \bar{x}_+)^2}{S_t} \right) \text{ e}$$

$$s.e.(\hat{\mu}_{x_0}) = \tilde{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x}_+)^2}{S_t}}.$$

Assim, temos,

$$IC(\mu_{x_0}, 1 - \alpha) : \hat{\mu}_{x_0} \pm t_{1-\alpha/2(n-2)} s.e.(\hat{\mu}_{x_0})$$

Observe que o menor IC ocorre quando  $x_0 = \bar{x}_+ \Rightarrow$  a predição (previsão) é mais precisa próxima do centro dos  $x_i$ 's e torna-se crescentemente imprecisa quando nos movemos para longe deste centro.

Em particular, não se deve tentar fazer previsões fora do intervalo de valores de  $x$  considerado na regressão, pois os dados não fornecem informação sobre a linearidade da relação entre  $Y$  e  $x$ .

Consideraremos agora o outro problema:

previsão de uma observação futura  $Y$  em  $x_0$ .

Este problema é fundamentalmente diferente do primeiro e envolve uma afirmação sobre uma observação futura  $y_{x_0}$  que ainda não foi observada.

Se realmente realizássemos um experimento para obter  $y_{x_0}$ , saberíamos o valor real de  $y_{x_0}$ , em contraste à questão da estimação de parâmetros onde nunca conheceremos o verdadeiro valor de um dado parâmetro.

Por outro lado, a medida que  $n$  cresce, somos capazes de estimar um dado parâmetro com precisão cada vez maior, ao passo que mesmo depois de acumular uma amostra muito grande, uma observação futura  $y_{x_0}$  nunca pode ser predita exatamente.

Observe que  $Y_{x_0}$  e  $\hat{\mu}_{x_0}$  são independentes pois supomos que  $Y_{x_0}$  é independente de  $Y_1, \dots, Y_n$ .

Assim,

$$Var(Y_{x_0} - \hat{\mu}_{x_0}) = \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x}_+)^2}{S_t} \right\}$$

Observe que  $Var(Y_{x_0} - \hat{\mu}_{x_0}) > \sigma^2$  devido ao fato de que mesmo se conhecêssemos  $\mu_{x_0}$  exatamente, não poderíamos prever  $Y_{x_0}$  com precisão melhor que  $\sigma^2$ , a variância de  $Y_{x_0}$ .

Assim,

$$Y_{x_0} \sim N \left( \mu_{x_0}, \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x}_+)^2}{S_t} \right\} \right)$$

Como  $\tilde{\sigma}_1^2$  é independente de  $Y_{x_0}$  e  $\hat{\mu}_{x_0}$  segue que

$$\frac{Y_{x_0} - \hat{\mu}_{x_0}}{s.e.(Y_{x_0})} \sim t_{(n-2)}$$

onde

$$s.e.(Y_{x_0}) = \tilde{\sigma}_1 \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x}_+)^2}{S_t} \right)^{1/2}$$

Assim,

$$IC(Y_{x_0}, 1 - \alpha) : \hat{\mu}_{x_0} \pm t_{1-\alpha/2(n-2)} s.e.(Y_{x_0})$$

que é o intervalo de predição.

Como  $s.e.(Y_{x_0}) > s.e.(\hat{\mu}_{x_0})$  e o mesmo estimador  $\tilde{\sigma}_1^2$  é usado em ambos os casos, o intervalo de predição sempre conterá o intervalo de confiança obtido anteriormente.

(EXEMPLO: SPINACH DATA)

### 4.5.2 Regressão Múltipla

A teoria de predição pode ser facilmente generalizada para o caso da regressão múltipla.

Considere o modelo usual:

$$Y_i = \mu_i + \epsilon_i, \quad \mu_i = \sum_j x_{ij}\beta_j$$

$i = 1, \dots, n$ . Seja  $x_0 = (x_{01}, \dots, x_{0k})^T$  um novo vetor de valores para as variáveis explicativas.

Queremos estimar  $\mu_{x_0} = \sum_j x_{0j}\beta_j = x_0^T\beta$  e

prever o valor  $y_{x_0}$  onde  $Y_{x_0} \sim N(\mu_{x_0}, \sigma^2)$ .

O estimador de  $\mu_{x_0}$  é  $\hat{\mu}_{x_0} = x_0^T\hat{\beta}$  e

$Var(\hat{\mu}_{x_0}) = x_0^T\{\sigma^2 C\}x_0 = \sigma^2 x_0^T C x_0$  onde  $C = (X^T X)^{-1}$ ,  $X = (x_{ij})$ .



Assim,

$s.e.(\hat{\mu}_{x_0}) = \tilde{\sigma}_1(x_0^T C x_0)^{1/2}$  onde  $\tilde{\sigma}_1^2$  é o estimador de  $\sigma^2$  sob o modelo considerado.

Como  $\frac{\hat{\mu}_{x_0} - \mu_{x_0}}{s.e.(\hat{\mu}_{x_0})} \sim t_{(n-k)}$  temos

$$IC(\mu_{x_0}, 1 - \alpha) : \hat{\mu}_{x_0} \pm t_{1-\alpha/2(n-k)} s.e.(\hat{\mu}_{x_0})$$

Os argumentos para a construção do intervalo de predição para  $Y_{x_0}$  também são imediatos.

Como  $Y_{x_0}$  é independente de  $Y_1, \dots, Y_n$  segue que  $Y_{x_0}$  é independente de  $\hat{\mu}_{x_0}$  tal que

$$Var(Y_{x_0} - \hat{\mu}_{x_0}) = \sigma^2(1 + x_0^T C x_0)$$

Além disso,  $\tilde{\sigma}_1^2$  é independente de  $Y_{x_0}$  e  $\hat{\mu}_{x_0}$  tal que

$$\frac{Y_{x_0} - \hat{\mu}_{x_0}}{s.e.(Y_{x_0})} \sim t_{(n-k)}$$

onde  $s.e.(Y_{x_0}) = \tilde{\sigma}_1(1 + x_0^T C x_0)^{1/2}$ . Assim,

$$IP(Y_{x_0}, 1 - \alpha) : \hat{\mu}_{x_0} \pm t_{1-\alpha/2(n-k)} s.e.(Y_{x_0})$$

## 4.6 Modelos Lineares Ponderados

Considere um modelo da forma

$$Y_i \stackrel{ind}{\sim} N(\mu_i, \frac{\sigma^2}{w_i}) \quad i = 1, \dots, n$$

onde  $w_1, \dots, w_n$  são números não-negativos conhecidos, chamados pesos.

Suponha que o vetor  $\mu = (\mu_1, \dots, \mu_n)^T$  siga algum modelo linear  $\mu \in L_1$ .

Junto com o modelo acima, temos um modelo que é conhecido como modelo linear ponderado com pesos  $w_1, \dots, w_n$ .

Observe que um peso nulo (implicando  $Var(Y) = \infty$ ) implica na eliminação da observação correspondente do modelo.

Assim, um conjunto de pesos cada um dos quais zero ou 1 pode ser usado para examinar um subconjunto dos dados.

Ponderações ocorrem naturalmente no caso onde as observações surgem na forma de médias baseadas em números de replicações desiguais. Por exemplo, os estimadores do modelo de análise de variância a um fator têm a forma

$\bar{Y}_{i+} \sim N(\mu_i, \frac{\sigma^2}{n_i})$  onde os  $n_i$ 's aparecem como pesos nas distribuições dos  $\bar{Y}_{i+}$ 's.

Ponderações são algumas vezes usadas de forma a eliminar parcialmente ou inflacionar a contribuição de uma dada observação ou subconjunto de observações.

O modelo linear ponderado é útil em diversos contextos e é fácil de ser analisado.

Considere uma forma mais geral do modelo linear ponderado  $Y_{n \times 1}$  tal que  $Y \sim N_n(\mu, \sigma^2 W^{-1})$  onde  $\mu \in L_1$ , onde  $W$  é uma matriz positiva definida.

Tal modelo é chamado modelo linear ponderado com matriz de pesos  $W$ . O primeiro caso analisado aqui corresponde a uma matriz  $W = \text{diag}\{w_1, \dots, w_n\}$ .

É fácil analisar este modelo usando os métodos já desenvolvidos. De fato, mostraremos que o MLP pode ser reduzido ao caso não ponderado.

Lembre que  $W$  é suposta ser conhecida enquanto que os parâmetros desconhecidos do modelo são  $\beta$  e  $\sigma^2$ .

Como  $W$  é positiva definida, existe  $A$ , não singular tal que  $W = A^T A$ . Então  $AY \sim N_n(A\mu, \sigma^2 I)$  pois  $Var(AY) = AW^{-1}A^T\sigma^2 = \sigma^2 I$ .

Assim,  $Y_0 = AY \sim N_n(\mu_o, \sigma^2 I)$  e encontramos que  $\mu \in L_1$  se e só se  $\mu_o \in L_{1'} = AL_1$ , este último sendo um espaço linear.

$\Rightarrow Y_0$  segue um ML padrão dado pelo espaço linear  $L_{1'}$  correspondente a uma nova matriz de modelo dada por  $X_0 = AX$ .

A partir deste resultado, podemos analisar o MLP usando os métodos de ML's já derivados.

Veremos brevemente que os resultados não exigem que calculemos a matriz  $A$  explicitamente, embora algumas vezes o seu conhecimento seja útil.

Uma possibilidade para obter  $A$ , se este for o caso, é via decomposição de Cholesky, onde  $A$  é escolhida como uma matriz triangular.

Observe também que podemos tomar

$A = \text{diag}\{w_1^{1/2}, \dots, w_n^{1/2}\}$  quando  $W$  é diagonal.

O estimador de  $\beta$  sob  $L_1$  ou  $L_1'$  é

$$\hat{\beta} = (X_0^T X_0)^{-1} X_0^T Y_0 = (X^T W X)^{-1} X^T W Y$$

que tem distribuição  $\hat{\beta} \sim N_k(\beta, \sigma^2 C)$  onde  $k = \dim \beta$  e  $C = (X^T W X)^{-1}$ .

Observe que a matriz “chapéu” para  $Y_0$  é

$H = X_0(X_0^T X_0)^{-1} X_0^T = A X (X^T W X)^{-1} X^T A^T$   
que exige, na verdade, o cálculo explícito de  $A$ .

O estimador para  $\sigma^2$  é

$$\begin{aligned}
 \tilde{\sigma}_1^2 &= \frac{1}{n-k} \|Y_0 - X_0 \hat{\beta}\|^2 = \frac{1}{n-k} \|AY - HAY\|^2 \\
 &= \frac{1}{n-k} (\|AY\|^2 - \|HAY\|^2) = \frac{1}{n-k} (Y^T A^T AY - Y^T A^T H^T H AY) \\
 &= \frac{1}{n-k} \{Y^T WY - Y^T W X (X^T W X)^{-1} X^T WY\}
 \end{aligned}$$

A distribuição de  $\tilde{\sigma}_1^2$  sob  $L_1$  é  $\sigma^2 \chi_{(f)}/f$  onde  $f = n - k$ .

Dado um subespaço de  $L_1$  na forma  $L_2 = \text{span}\{x_1, \dots, x_l\}$  onde  $x_1, \dots, x_k$  são as colunas de  $X$ , encontramos que o subespaço linear de  $L_1$ , correspondente é  $L_{2'} = \text{span}\{Ax_1, \dots, Ax_l\}$ .

Assim, podemos facilmente derivar a forma do teste  $F$  para testar  $L_2$  sob  $L_1$ .

Seguindo a notação anterior, o quadrado da norma no numerador do teste  $F$  é

$$\begin{aligned} D &= \|AX_1(\hat{\beta}^{(1)} - \hat{\psi}) + AX_2\hat{\beta}^{(2)}\|^2 = \\ &= (\hat{\beta}^{(1)} - \hat{\psi})^T X_1^T W X_1 (\hat{\beta}^{(1)} - \hat{\psi}) + \hat{\beta}^{(2)T} X_2^T W X_2 \hat{\beta}^{(2)} + \\ &\quad + 2(\hat{\beta}^{(1)} - \hat{\psi})^T X_1^T W X_2 \hat{\beta}^{(2)} \end{aligned}$$

Sob  $L_2$ ,

$$F = \frac{D/(k-l)}{\tilde{\sigma}_1^2} \sim F_{(k-l, n-k)}.$$