Orthogonal polynomials

Stat 526
5-6-25

①

Suppose that a factor has $\underline{a}$ levels,
so it has $a-1$ df.

For the same df, we could fit an $\underline{a-1}$-order polynomial.

Let the levels of the factor be $x_1, x_2, \ldots, x_a$

$i = 1, \ldots, a \qquad j = 1, \ldots, n \qquad N = an$

Let $P_0(x) = 1$

②

Let $P_1(x) = a_0 + a_1 x$

But we want $P_0(x) \perp P_1(x)$,

i.e. $\sum_i \underbrace{P_0(x_i)}_{1} P_1(x_i) = 0$

$$\sum_{i=1}^{a} (a_0 + a_1 x_i) = a\, a_0 + a_1 \sum_{i=1}^{a} x_i$$

$$a_0 = -\frac{a_1 \sum x_i}{a} = -a_1 \bar{x}$$

$\leftarrow$ arbitrary

So $P_1(x) = -a_1 \bar{x} + a_1 x = a_1 (x - \bar{x})$

③

let $P_2(x) = b_0 + b_1 x + b_2 x^2$

Need $\sum_{i=1}^{a} P_0(x_i) P_2(x_i) = 0$

and $\sum_{i=1}^{a} P_1(x_i) P_2(x_i) = 0$

But it's easier to get a solution if we write $P_2(x) = c_0 + c_1(x-\bar{x}) + c_2(x-\bar{x})^2$

④

$$\sum_{i=1}^{a} \left[c_0 + c_1(x-\bar{x}) + c_2(x-\bar{x})^2\right] = 0$$

$$a\,c_0 + c_1 \underbrace{\sum_{i=1}^{a}(x-\bar{x})}_{S_1} + c_2 \underbrace{\sum_{i=1}^{a}(x-\bar{x})^2}_{S_2}$$

But $S_1 = 0$

$$a\,c_0 + c_2 S_2 = 0$$

$$c_0 = -\frac{c_2}{a} S_2$$

(5)

$$\sum_{i=1}^{a} (x-\bar{x})\left[c_0 + c_1(x-\bar{x}) + c_2(x-\bar{x})^2\right] = 0$$

$$c_0 \cancel{S_1} + c_1 S_2 + c_2 S_3 = 0$$

$$c_1 = -c_2 \frac{S_3}{S_2}$$

$$P_2(x) = -\frac{c_2}{a} S_2 + \frac{-c_2 S_3}{S_2}(x-\bar{x}) + c_2 (x-\bar{x})^2$$

arbitrary

$$= c_2\left[-\frac{1}{a}S_2 - \frac{S_3}{S_2}(x-\bar{x}) + (x-\bar{x})^2\right]$$

(6)

<u>Special case</u>

Suppose that the levels of the factor are equally spaced.

Then $\quad x_i = x_1 + (i-1)d$

So $\quad \bar{x} = \frac{1}{a} \sum_{i=1}^{a} \left(x_1 + (i-1)d\right)$

$$= x_1 + \frac{d}{a} \underbrace{\sum_{i=1}^{a} (i-1)}_{\sum_{k=0}^{a-1} k \, = \, \frac{(a-1)a}{2}}$$

(7)

$$\bar{x} = x_1 + \frac{d}{a}\frac{(a-1)a}{2} = x_1 + \frac{d}{2}(a-1)$$

$$x_i - \bar{x} = [x_1 + (i-1)d] - [x_1 + \tfrac{d}{2}(a-1)]$$

$$= d\left(i - 1 - \frac{a-1}{2}\right)$$

$$= d\left(i - \frac{a+1}{2}\right)$$

$$S_1 = \sum_{i=1}^{a}(x_i - \bar{x}) = \sum_{i=1}^{a} d\left(i - \frac{a+1}{2}\right)$$

$$= d\left[\frac{a(a+1)}{2} - \frac{a(a+1)}{2}\right] = 0$$ (we knew this already)

(8)

$$S_2 = \sum_{i=1}^{a}(x_i - \bar{x})^2 = \sum_{i=1}^{a} d^2\left(i - \frac{a+1}{2}\right)^2$$

$$= d^2 \sum_{i=1}^{a}\left[i^2 - (a+1)i + \frac{(a+1)^2}{4}\right]$$

$$= d^2\left[\frac{a(a+1)(2a+1)}{6} - (a+1)\frac{a(a+1)}{2} + a\frac{(a+1)^2}{4}\right]$$

$$= \frac{d^2}{12}a(a^2-1)$$

$$S_3 = \sum_{i=1}^{a} d^3\left(i - \frac{a+1}{2}\right)^3 = 0$$ (only works because the x-values were equally spaced)

Note: $S_5 = S_7 = \ldots$ will also be 0 (9)

What do the orthogonal polynomials look like, now that the x-values are equally spaced?

$$P_0(x) = 1$$

$$P_1(x) = a_1(x-\bar{x}) = a_1 \frac{d(x-\bar{x})}{d} = \lambda_1\left(\frac{x-\bar{x}}{d}\right)$$

$$P_2(x) = c_2\left[-\frac{S_2}{a} - \frac{S_3}{S_2}(x-\bar{x}) + (x-\bar{x})^2\right]$$

(10)

$$= c_2\left[-\frac{1}{\cancel{a}}\frac{d^2}{12}\cancel{a}(a^2-1) - 0 + (x-\bar{x})^2\right]$$

$$= c_2 d^2\left[-\frac{1}{12}(a^2-1) + \left(\frac{x-\bar{x}}{d}\right)^2\right]$$

$$P_2(x) = \lambda_2\left[\left(\frac{x-\bar{x}}{d}\right)^2 - \frac{a^2-1}{12}\right]$$

Numerical example:

A factor has 5 levels

$x_1 = 10$
$x_2 = 20$
$x_3 = 30$
$x_4 = 40$
$x_5 = 50$

(11)

$P_0(x) = 1$ $\qquad \bar{x} = 30$

$P_1(x) = \lambda_1\left(\frac{x-\bar{x}}{10}\right)$

$P_2(x) = \lambda_2\left[\left(\frac{x-\bar{x}}{10}\right)^2 - 2\right]$

We won't finish by finding $P_3(x)$ + $P_4(x)$

$$\begin{bmatrix} 1 & -2 & 2 & \cdot & \cdot \\ 1 & -1 & -1 & \cdot & \cdot \\ 1 & 0 & -2 & \cdot & \cdot \\ 1 & 1 & -1 & \cdot & \cdot \\ 1 & 2 & 2 & \cdot & \cdot \end{bmatrix}$$

Design matrix, with orthogonal columns

(12)

HW #5

15.11
15.15 (ANCOVA)

15.11. A soft drink distributor is studying the effectiveness of delivery methods. Three different types of hand trucks have been developed, and an experiment is performed in the company's methods engineering laboratory. The variable of interest is the delivery time in minutes (y); however, delivery time is also strongly related to the case volume delivered (x). Each hand truck is used four times and the data that follow are obtained. Analyze these data and draw appropriate conclusions. Use $\alpha = 0.05$.

Hand Truck Type					
1		**2**		**3**	
y	x	y	x	y	x
27	24	25	26	40	38
44	40	35	32	22	26
33	35	46	42	53	50
41	40	26	25	18	20

15.15. Four different formulations of an industrial glue are being tested. The tensile strength of the glue when it is applied to join parts is also related to the application thickness. Five observations on strength (y) in pounds and thickness (x) in 0.01 inches are obtained for each formulation. The data are shown in the following table. Analyze these data and draw appropriate conclusions.

Glue Formulation							
1		**2**		**3**		**4**	
y	x	y	x	y	x	y	x
46.5	13	48.7	12	46.3	15	44.7	16
45.9	14	49.0	10	47.1	14	43.0	15
49.8	12	50.1	11	48.9	11	51.0	10
46.1	12	48.5	12	48.2	11	48.1	12
44.3	14	45.2	14	50.3	10	48.6	11