Stat 566
5-7-15

Orthogonal polynomials

①

Suppose that a factor has a levels,

& so $a-1$ df

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

Let the levels of the factor be $x_1, \dots, x_a$

Model:

$$y_{ij} = \beta_0 P_0(x_i) + \beta_1 P_1(x_i) + \dots + \beta_{a-1} P_{a-1}(x_i) + \varepsilon_{ij}$$

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

②

Let $P_0(x) = 1$

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

For these two polynomials to be $\perp$, we need

$$\sum_i P_0(x_i) P_1(x_i) = 0$$

$$= \sum_i (a_0 + a_1 x_i) = a\,a_0 + a_1 \sum x_i$$

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

$$= -a_1 \bar{x}$$

③

$$P_1(x) = a_0 + a_1 x = -a_1\bar{x} + a_1 x$$
$$= a_1(x-\bar{x})$$

↑ arbitrary

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

Need $\sum_i P_0(x_i) P_2(x_i) = 0$

and $\sum_i P_1(x_i) P_2(x_i) = 0$

④

$$\sum_i P_2(x_i) = 0$$
$$\sum_i (b_0 + b_1 x_i + b_2 x_i^2)$$

and $\sum_i (x_i - \bar{x})(b_0 + b_1 x_i + b_2 x_i^2) = 0$

Simpler derivation:

Let $P_2(x) = c_0 + c_1(x-\bar{x}) + c_2(x-\bar{x})^2$

(5)

Now, $$\sum_i \left[c_0 + c_1(x_i - \bar{x}) + c_2(x_i - \bar{x})^2 \right] = 0$$

$$a c_0 + c_1 \cancel{S_1} + c_2 S_2 = 0 \quad \text{but } S_1 = 0$$

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

and $$\sum_i (x_i - \bar{x})\left(c_0 + c_1(x_i - \bar{x}) + c_2(x_i - \bar{x})^2\right) = 0$$

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

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

(6)

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

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

arbitrary

and so forth

Special case

Suppose that the levels of the factor are equally spaced.

7

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

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

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

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

8

$$x_i - \bar{x} = \cancel{x_1} + (i-1)d - \left(\cancel{x_1} + \frac{d}{2}(a-1)\right)$$

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

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

$$S_1 = \sum_{i=1}^{a} d\left(i - \frac{a+1}{2}\right) = 0$$ (true even when the x_i terms are not evenly spaced)

$$S_2 = \sum_{i=1}^{a} d^2\left(i - \frac{a+1}{2}\right)^2 = d^2\left[\sum_i i^2 - (a+1)\sum_i i + \frac{(a+1)^2}{4}a\right]$$

$$= d^2\left[\frac{a(a+1)(2a+1)}{6} - (a+1)\frac{a(a+1)}{2} + \frac{a(a+1)^2}{4}\right]$$ (9)

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

$$S_3 = \sum_{i=1}^{a} d^3\left(i - \frac{a+1}{2}\right)^3 = 0$$ (only true in the evenly-spaced case)

Also, $S_5 = S_7 =$ all other odd S terms $= 0$

(10)

$$P_0(x) = 1$$

$$P_1(x) = a_1(x-\bar{x}) = a_1 d\,\frac{(x-\bar{x})}{d}$$

$$= \lambda_1 \frac{x-\bar{x}}{d}$$

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

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

$$= c_2 d^2\left[\left(\frac{x-\bar{x}}{d}\right)^2 - \frac{a^2-1}{12}\right] = \lambda_2\left[\left(\frac{x-\bar{x}}{d}\right)^2 - \frac{a^2-1}{12}\right]$$

(11)

Actual example: A factor as 5 levels, and $n=1$

$P_0(x) = 1$

$P_1(x) = \lambda_1 \dfrac{x-\bar{x}}{10} \quad (\lambda_1 = 1)$

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

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

$d = 10$

$\bar{x} = 30$

Design matrix:

$$\begin{array}{ccccc} & & & P_3 & P_4 \\ \end{array}$$
$$\begin{bmatrix} 1 & -2 & 2 & \cdot & \vdots \\ 1 & -1 & -1 & \cdot & \cdot \\ 1 & 0 & -2 & \cdot & \cdot \\ 1 & 1 & -1 & \cdot & \cdot \\ 1 & 2 & 2 & \cdot & \cdot \end{bmatrix}$$

(12)

Your textbook has a table of orthogonal polynomials

HW #5 due 5/14

15.11
15.15

ANCOVA

15.9. Problem 12.10 suggests using ln (s^2) as the response [refer to part (b)]. Does the Box–Cox method indicate that a transformation is appropriate?

15.10. Myers, Montgomery, Vining and Robinson (2010) describe an experiment to study spermatozoa survival. The design factors are the amount of sodium citrate, the amount of glycerol, and equilibrium time, each at two levels. The response variable is the number of spermatozoa that survive out of 50 that were tested at each set of conditions. The data are shown in the following table:

Sodium Citrate	Glycerol	Equilibrium Time	Number Survived
−	−	−	34
+	−	−	20
−	+	−	8
+	+	−	21
−	−	+	30
+	−	+	20
−	+	+	10
+	+	+	25

Analyze the data from this experiment with logistic regression.

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

15.12. Compute the adjusted treatment means and the standard errors of the adjusted treatment means for the data in Problem 15.11.

15.13. The sums of squares and products for a single-factor analysis of covariance follow. Complete the analysis and draw appropriate conclusions. Use $\alpha = 0.05$.

Source of Variation	Degrees of Freedom	Sums of Squares and Products		
		x	xy	y
Treatment	3	1500	1000	650
Error	12	6000	1200	550
Total	15	7500	2200	1200

15.14. Find the standard errors of the adjusted treatment means in Example 15.5.

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

15.16. Compute the adjusted treatment means and their standard errors using the data in Problem 15.15.

15.17. An engineer is studying the effect of cutting speed on the rate of metal removal in a machining operation. However, the rate of metal removal is also related to the hardness of the test specimen. Five observations are taken at each cutting speed. The amount of metal removed (y) and the hardness of the specimen (x) are shown in the following table. Analyze the data using an analysis of covariance. Use $\alpha = 0.05$.

Cutting Speed (rpm)					
1000		1200		1400	
y	x	y	x	y	x
68	120	112	165	118	175
90	140	94	140	82	132
98	150	65	120	73	124
77	125	74	125	92	141
88	136	85	133	80	130

15.18. Show that in a single-factor analysis of covariance with a single covariate a $100(1 - \alpha)$ percent confidence interval on the ith adjusted treatment mean is

$$\bar{y}_{i.} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \pm t_{\alpha/2,a(n-1)-1}$$