Ortogonálny plán pre dvojfaktorový lineárny regresný model

Majme nameranú závislosť $y = f(t, \upsilon)$ n=2

Určite odhad $\hat{y} = \hat{\theta}'_0 + \hat{\theta}'_1 t + \hat{\theta}'_2 \upsilon$ ak sa premenné menia v rozsahoch:

$$3\,\text{hod} \le t \le 5\,\text{hod} \qquad 210\ ^\circ\text{C} \le \upsilon \le 230\ ^\circ\text{C}$$

Počet experimentálnych bodov je: $B = 2^n = 4$, počet odhadovaných parametrov je $n + 1 = 3$. Budeme uvažovať $q = 2$ merania v každom bode.

Vstupy $u'_i = t, \upsilon$ pretransformujeme na bezrozmerné premenné $u_i \in \langle -1, 1 \rangle$

$$u_i = \frac{u'_i - \dfrac{u'_{i\max} + u'_{i\min}}{2}}{\dfrac{u'_{i\max} - u'_{i\min}}{2}} = \frac{2u'_i - u'_{i\max} - u'_{i\min}}{u'_{i\max} - u'_{i\min}}$$

$$u_1 = \frac{t - \dfrac{5+3}{2}}{\dfrac{5-3}{2}} = t - 4 \qquad t = u_1 + 4$$

$$u_2 = \frac{\upsilon - \dfrac{230+210}{2}}{\dfrac{230-210}{2}} = 0.1\upsilon - 22 \qquad \upsilon = 10u_2 + 220$$

Po dosadení nových premenných za pôvodné premenné bude:

$$\hat{y} = \hat{\theta}'_0 + \hat{\theta}'_1(u_1 + 4) + \hat{\theta}'_2(10u_2 + 220) = \hat{\theta}_0 + \hat{\theta}_1 u_1 + \hat{\theta}_2 u_2$$

i	t	υ	u_0	u_1	u_2	y_1	y_2	$\bar{y}$	$\hat{y}$
1	5	230	1	1	1	82.6	82.7	82.65	83.65
2	3	230	1	-1	1	79.3	79.1	79.20	78.2
3	5	210	1	1	-1	89.6	89.6	89.60	88.60
4	3	210	1	-1	-1	82.2	82.1	82.15	83.15

Podľa vzťahu $\hat{\theta}_j = \dfrac{1}{2^n}\sum_{i=1}^{2^n} \bar{y}_i u_{ij} = \dfrac{1}{2^n}\sum_{i=1}^{2^n} \bar{y}_i (-1)^{\mathrm{int}\left(\frac{i-1}{2^{j-1}}\right)}$ určíme hodnoty parametrov:

$$\hat{\theta}_0 = \frac{\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4}{4} = 83.4$$

$$\hat{\theta}_1 = \frac{\bar{y}_1 - \bar{y}_2 + \bar{y}_3 - \bar{y}_4}{4} = 2.725$$

$$\hat{\theta}_2 = \frac{\overline{y}_1 + \overline{y}_2 - \overline{y}_3 - \overline{y}_4}{4} = 2.475$$

Otestujeme štatistickú významnosť parametrov:

$$Q_{\bar{y}} = \sum_{i=1}^{4} \sum_{j=1}^{2} \left(Y_{ij} - \overline{Y}_i\right)^2 = 2\left(0.05^2 + 0.1^2 + 0^2 + 0.05^2\right) = 0.03$$

$$\nu_{\bar{Y}} = 2^n \times (q-1) = 4 \qquad S^2_{\bar{Y}} = \frac{Q_{\bar{Y}}}{\nu_{\bar{Y}}} = \frac{0.03}{4} = 0.0075$$

$$S^2_{\hat{\theta}_i} = \frac{S^2_{\bar{Y}}}{2^n} = \frac{0.0075}{4} = 0.001875$$

Ak je $\alpha = 0.01$ potom $t_{0.995}(4) = 4.604$ a

$$\left|\hat{\theta}_i\right| > \delta = S_{\hat{\theta}_i}.t_{0.995}(4) = 4.604\sqrt{0.001875} = 0.199$$

Parametre sú štatisticky významné. Otestujeme teraz štruktúru modelu:

$$Q_{\tilde{Y}} = \sum_{I=1}^{4} 2\left(\overline{Y}_i - \hat{Y}_i\right)^2 = 2\left(1^2 + 1^2 + 1^2 + 1^2\right) = 8$$

$$\upsilon_{\tilde{y}} = 2^n - n - 1 = 1 \qquad S^2_{\tilde{y}} = \frac{Q_{\tilde{Y}}}{\nu_{\tilde{Y}}} = 8$$

Pre $\alpha = 0.01$ je $w_{1-\alpha}(\nu_{\tilde{Y}}, \nu_{\bar{Y}}) = w_{0.99}(1,4) = 21.2$

$$\frac{S^2_{\tilde{Y}}}{S^2_{\bar{Y}}} = \frac{8}{0.0075} = 1066.67 > 21.2,$$

t.j. lineárny model nevyhovuje

$$\hat{y} = 83.4 + 2.725(t - 4) - 2.475(0.1\upsilon - 22) = 126.95 + 2.275t - 0.2475\upsilon$$

Boxov ortogonálny kompozičný kvadratický 2-faktorový plán

Počet regresných koeficientov $\frac{1}{2}(n+1)(n+2)=6$

Počet bodov: $N=2^n+2n+1=2^2+2.2+1=9$ $q=2$

$$\mu=\sqrt{\frac{2^n}{N}}=\sqrt{\frac{4}{9}}=\frac{2}{3} \qquad \alpha=\pm\sqrt{\frac{1}{2}\mu N(1-\mu)}=\pm\sqrt{\frac{1}{2}\frac{2}{3}9\left(1-\frac{2}{3}\right)}=1$$

$$\hat{y}=\hat{\theta}_0+\hat{\theta}_1u_1+\hat{\theta}_2u_2+\hat{\theta}_3u_1^2+\hat{\theta}_4u_2^2+\hat{\theta}_5u_1u_2$$

i	u_0	u_1	u_2	$u_1^2-\mu$	$u_2^2-\mu$	u_1u_2	y_1	y_2	$\overline{y}$	$\widehat{y}$
1	1	1	1	1/3	1/3	1	82,6	82,7	82,65	82,567
2	1	-1	1	1/3	1/3	-1	79,3	79,1	79,20	79,133
3	1	1	-1	1/3	1/3	-1	89,6	89,6	89,60	89,600
4	1	-1	-1	1/3	1/3	1	82,2	82,1	82,15	82,167
5	1	1	0	1/3	-2/3	0	89,2	89,1	89,15	89,233
6	1	-1	0	1/3	-2/3	0	83,7	83,8	83,75	83,800
7	1	0	1	-2/3	1/3	0	82,2	82,1	82,15	82,300
8	1	0	-1	-2/3	1/3	0	87,3	87,4	87,35	87,333
9	1	0	0	-2/3	-2/3	0	88,0	88,2	88,10	87,966

Disperzná matica

$$p_3=\frac{1}{\mu^2N}=\frac{1}{4} \quad p_2=\frac{2}{\mu^2(1-\mu)^2N^2}=\frac{1}{2} \quad p_1=\frac{1}{\mu N}=\frac{1}{6} \quad p_0=\frac{1}{N}=\frac{1}{9} \quad p_0+\mu^2p_2n=\frac{5}{9}$$

Odhad parametrov

$$\hat{\theta}_j^*=p_1\sum_{i=1}^{N}\overline{y}_iu_{ij} \qquad j=1,2,\cdots,n$$

$$\hat{\theta}_j^*=p_2\sum_{i=1}^{N}\overline{y}_i\left[u_{i,j-n}^2-\mu\right] \qquad j=n+1,\dots,2n$$

$$\hat{\theta}_j^*=p_3\sum_{i=1}^{N}\overline{y}_iu_{ix}u_{iz} \qquad j=\frac{(x+1)(2n-z)}{2}+z=2n+1,\cdots,k\ ,\ 1\le x<z\le n$$

$$\hat{\theta}_0^*=p_0\sum_{i=1}^{n}\overline{y}_i-\mu\sum_{i=1}^{n}\hat{\theta}_{i+n}^*$$

$$\hat{\theta}_1=p_1\sum_{i=1}^{N}\overline{y}_iu_{i1}=\frac{1}{6}\left[\overline{y}_1-\overline{y}_2+\overline{y}_3-\overline{y}_4+\overline{y}_5-\overline{y}_6\right]=2.717$$

$$\hat{\theta}_2 = p_1\sum_{i=1}^{N}\bar{y}_i u_{i2} = \frac{1}{6}\left[\bar{y}_1 + \bar{y}_2 - \bar{y}_3 - \bar{y}_4 + \bar{y}_7 - \bar{y}_8\right] = -2.517$$

$$\hat{\theta}_3 = p_2\sum_{i=1}^{N}\bar{y}_i\left[u_{i1}^2 - \mu\right] = \frac{1}{2}\left[\frac{1}{3}\left(\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_5 + \bar{y}_6\right) - \frac{2}{3}\left(\bar{y}_7 + \bar{y}_8 + \bar{y}_9\right)\right] = -1.45$$

$$\hat{\theta}_4 = p_2\sum_{i=1}^{N}\bar{y}_i\left[u_{i2}^2 - \mu\right] = \frac{1}{2}\left[\frac{1}{3}\left(\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_7 + \bar{y}_8\right) - \frac{2}{3}\left(\bar{y}_5 + \bar{y}_6 + \bar{y}_9\right)\right] = -3.15$$

$$\hat{\theta}_5 = p_3\sum_{i=1}^{N}\bar{y}_i u_{i1}u_{i2} = \frac{1}{4}\left[\bar{y}_1 - \bar{y}_2 - \bar{y}_3 + \bar{y}_4\right] = -1$$

$$\hat{\theta}_0 = p_0\sum_{i=1}^{n}\bar{y}_i - \mu\sum_{i=1}^{n}\hat{\theta}_{i+n} =$$

$$= \frac{1}{9}\left[\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_5 + \bar{y}_6 + \bar{y}_7 + \bar{y}_8 + \bar{y}_9\right] - \frac{2}{3}\left(\hat{\theta}_3 + \hat{\theta}_4\right) = 87.967$$

Testovanie štatistickej významnosti regresných koeficientov: $\alpha = 1\%$

Výberový rozptyl vzoriek voči priemeru skupiny (odhad disperzie šumov)

$$Q_{\bar{y}} = \sum_{i=1}^{N}\sum_{j=1}^{q}\left(y_{ij} - \bar{y}_i\right)^2 = \sum_{i=1}^{9}\sum_{j=1}^{2}\left(y_{ij} - \bar{y}_i\right)^2 =$$

$$= 2\left[0.05^2 + 0.1^2 + 0^2 + 0.05^2 + 0.05^2 + 0.05^2 + 0.05^2 + 0.05^2 + 0.1^2\right] = 0.07$$

$$\upsilon_{\bar{y}} = N(q-1) = 9 \qquad S_{\bar{y}}^2 = \frac{0.07}{9} = 0.0078$$

$$t_{1-\alpha}\left(\nu_{\bar{y}}\right) = t_{0.995}(9) = 3.25$$

$$\delta_i = S_{\hat{\theta}_i} t_{0.995}(9) = 3.25\, S_{\hat{\theta}_i} \quad i = \{o, z, k, l\}$$

$$S_{\theta_0}^2 = \left(p_0 + \mu^2 n p_2\right)S_{\bar{y}}^2 \qquad \delta_{\hat{\theta}_0} < \left|\hat{\theta}_0\right|$$

$$S_{\theta_i}^2 = p_1 S_{\bar{y}}^2 \quad i = 1, 2, \ldots, n \quad \delta_{\hat{\theta}_i} < \left\{\left|\hat{\theta}_1\right|, \left|\hat{\theta}_2\right|\right\}$$

$$S_{\theta_i}^2 = p_2 S_{\bar{y}}^2 \quad i = n+1, \ldots, 2n \quad \delta_{\hat{\theta}_i} < \left\{\left|\hat{\theta}_3\right|, \left|\hat{\theta}_4\right|\right\}$$

$$S_{\theta_i}^2 = p_3 S_{\bar{y}}^2 \quad i = 2n+1, \ldots, k \quad \delta_{\hat{\theta}_i} < \left|\hat{\theta}_5\right|$$

Koeficient θ_j z modelu vypustíme, ak

$$-\delta_j \le \hat{\theta}_j^* \le \delta_j$$

$$\left|\hat{\theta}_0\right| = 87.967 \ge 3.25\sqrt{\frac{5}{9}0.0078} = 3.25 * 0.0657 = 0.214$$

$$\left|\hat{\theta}_1\right| = 2.717 \ge 3.25\sqrt{\frac{1}{6}0.0078} = 3.25 * 0.0360 = 0.117$$

$$\left|\hat{\theta}_2\right| = 2.517 \ge 3.25\sqrt{\frac{1}{6}0.0078} = 3.25 * 0.0360 = 0.117$$

$$\left|\hat{\theta}_3\right| = 1.45 \geq 3.25\sqrt{\frac{1}{2}0.0078} = 3.25 * 0.0624 = 0.209$$

$$\left|\hat{\theta}_4\right| = 3.15 \geq 3.25\sqrt{\frac{1}{2}0.0078} = 3.25 * 0.0624 = 0.209$$

$$\left|\hat{\theta}_5\right| = 1 \geq 3.25\sqrt{\frac{1}{4}0.0078} = 3.25 * 0.0441 = 0.143$$

Keďže všetky koeficienty sú štatisticky významné, môžeme vypočítať $\hat{y}$:

$$\hat{y} = 87.967 + 2.717u_1 - 2.517u_2 - 1.45u_1^2 - 3.15u_2^2 - u_1u_2$$

Testovanie adekvátnosti kvadratického modelu: $\alpha = 1\%$

$$Q_{\tilde{y}} = q\sum_{i=1}^{N}\left(\bar{y}_i - \hat{y}_i\right)^2 = 2\sum_{i=1}^{9}\left(\bar{y}_i - \hat{y}_i\right)^2 =$$

$$= 2\left[0.083^2 + 0.067^2 + 0^2 + 0.017^2 + 0.083^2 + 0.05^2 + 0.15^2 + 0.017^2 + 0.134^2\right] = 0.123$$

$$\nu_{\tilde{Y}} = N - k - 1 = 9 - 6 = 3$$

$$S_{\tilde{Y}}^2 = \frac{0.123}{3} = 0.0412$$

$$w_{1-\alpha}(\nu_{\tilde{Y}}, \nu_{\bar{Y}}) = w_{0.99}(3,9) = 6.99$$

$$\frac{S_{\tilde{Y}}^2}{S_{\bar{Y}}^2} = \frac{0.0412}{0.0078} = 5.28 \leq 6.99 \qquad \rightarrow \qquad \text{kvadratický model vyhovuje}$$

$$\hat{y} = 87.967 + 2.717(t-4) - 2.517(0.1\vartheta - 22) - 1.45(t-4)^2 - 3.15(0.1\vartheta - 22)^2 - (t-4)(0.1\vartheta - 22) =$$
$$= -1500.427 + 36.317t + 14.008\vartheta - 1.45t^2 - 0.032\vartheta^2 - 0.1t\vartheta$$

$$\left.\begin{array}{l} \dfrac{\partial}{\partial t}\hat{y} = 0 \\ \dfrac{\partial}{\partial \vartheta}\hat{y} = 0 \end{array}\right\} \Rightarrow t^* = 5.26\text{hod}, \quad \vartheta = 210.66°\text{C} \quad \text{a} \quad \hat{y}^* = 89.87$$