SIGNIFIKANTNOST REGRESNÍCH KOEFICIENTŮ URČENÝCH METODOU LWS.

PAVEL PLÁT

plat@klinux.fjfi.cvut.cz
Katedra matematiky, ČVUT – FJFI, Praha

ABSTRACT

Tento příspěvek navazuje na dosud publikované výsledky o LWS a zabývá testováním signifikantnosti jednotlivých koeficientů a celého modelu určeného pomocí LWS.

NEJMENŠÍ VÁŽENÉ ČTVERCE (LWS)

Uvažujme lineární regresní model, který můžeme zapsat ve tvaru

LINEÁRNÍ REGRESNÍ MODEL

$$Y = X\beta + e,$$

kde Y je vysvětlovaná proměnná, X je matice regresorů a e je vektor náhodných fluktuací. Při konstrukci odvozených statistik předpokládáme, náhodné veličiny e_i, $i \in N$ jsou nezávislé a normálně rozdělené. Dále pro nejmenší vážené čtverce uvažujeme po částech lineární váhovou funkci w.

PO ČÁSTECH LINEÁRNÍ VÁHOVÁ FUNKCE

$$w(x) = -k_r x + c_r, \quad pro\ x \in [a_{r-1}, a_r),$$

kde $0 = a_0 < a_1 < \ldots < a_m = 1$ a $k_r \geq 0$, $r \in \{1, 2, \ldots, m\}$ jsou parametry. Volbou těchto parametrů můžeme měnit vlastnosti odhadu. Čísla c_r, $r \in \{1, 2, \ldots, m\}$ musí být určeny tak, aby funkce w byla spojitá.

Odhad regresních koeficientů metodou nejmenších vážených čtverců

NEJMENŠÍ VÁŽENÉ ČTVERCE (LWS)

$$\hat{\beta}^{(LWS,n,w)} = \arg\min_{\beta \in R} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta),$$

$$\mathcal{L}\left(\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)\right) \to \mathcal{N}\left(0\ ,\ V\left(\hat{\beta}^{(LWS,n,w)}\right)\right).$$

je $\sqrt{n}-$konzistentní, asymptoticky normální a pro odhad kovarianční matice můžeme použít následující statistiku

ODHAD KOVARIANČNÍ MATICE

$$\hat{V}_n = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} w^2\left(\frac{i-1}{n}\right) r_{(i)}^2(\hat{\beta}^{(LWS,n,w)})}{\left(\sum_{r=1}^{m-1} k_r \left(\frac{a_{r+1}^2 - a_r^2}{2} - \frac{1}{\pi}\left(e^{-\left(\Phi^{-1}\left(\frac{1+a_r}{2}\right)\right)^2} - e^{-\left(\Phi^{-1}\left(\frac{1+a_r+1}{2}\right)\right)^2}\right)\right)\right)^2} \cdot \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i^T\right)^{-1},$$

$$\hat{V}_n \xrightarrow{P} V\left(\hat{\beta}^{(LWS,n,w)}\right).$$

KOEFICIENT DETERMINACE, TEST SIGNIFIKANCE KOEFICIENTŮ

Pro otestování signifikance jednotlivých vysvětlujících proměnných použijeme standardizované odhady regresních koeficientů, které porovnáme oproti Studentovu t s $n-p$ stupni volnosti.

STANDARDIZOVANÉ ODHADY KOEFICIENTŮ

$$\hat{t}_i(w) = (\hat{V}_n)_{ii}^{-\frac{1}{2}} \hat{\beta}_i^{(LWS,n,w)}.$$

Dále definujme čísla k_i vztahem $r_{k_i}^2(\hat{\beta}^{(LWS,n,w)}) = r_{(i)}^2(\hat{\beta}^{(LWS,n,w)})$, položme $R_{W0}^2 = \sum\limits_{i=1}^{n} w^2\left(\frac{i-1}{n}\right)(Y_{k_i} - \bar{Y})^2$ (respektive $R_{W0}^2 = \sum\limits_{i=1}^{n} w^2\left(\frac{i-1}{n}\right) Y_{k_i}^2$ pro model bez interceptu) a $S_{WR}^2 = \sum\limits_{i=1}^{n} w^2\left(\frac{i-1}{n}\right) r_{(i)}^2(\hat{\beta}^{(LWS,n,w)})$. Koeficientem determinace založeným na vážených reziduích pak rozumíme

KOEFICIENT DETERMINACE

$$R_W^2 = \frac{R_{W0}^2 - S_{WR}^2}{R_{W0}^2}.$$

PŘÍKLAD

X_1, X_2 nezávislé s rovnoměrným rozdělením $U_{[1,20]}$, Y splňuje model

$$Y = 1 + 2X_1 + e,$$

kde $\mathcal{L}(e_i) = \mathcal{N}(0, 5)$. Model je tedy nezávislý na veličině X_2. Hodnota Y_{30} byla změněna – představuje chybu v datech. Pro LWS je použita váhová funkce s parametry $m = 3$, $a_1 = 0.5502$, $a_2 = 0.95$, $k_1 = 0$, $k_2 = 2.5$, $k_3 = 0.01$.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X_1	9.08	19.84	12.02	19.31	9.90	13.12	8.76	18.33	13.20	12.09	18.20	2.08	18.52	12.37	7.73
X_2	8.28	16.68	11.40	10.31	17.31	9.02	13.93	10.94	1.31	1.52	1.49	1.80	9.68	3.49	12.97
Y	24.74	44.11	23.01	35.65	20.68	23.36	13.43	39.22	31.23	25.93	39.65	5.36	39.60	26.89	17.28
i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
X_1	14.71	14.50	1.61	13.51	3.21	4.08	15.37	7.37	7.33	4.52	5.19	2.15	17.52	16.15	6.63
X_2	7.50	16.17	5.72	19.02	4.43	12.43	16.35	12.29	10.91	14.04	8.11	3.97	19.32	16.04	15.98
Y	26.18	27.29	1.56	28.11	11.91	8.72	28.34	23.13	15.45	8.73	11.68	12.45	39.28	43.85	150.00

VÝSLEDKY LS

Model se zahrnutím X_1, X_2

β_0	t_0	p	β_1	t_1	p	β_2	t_2	p	R^2
-13.96	-1.42	0.17	3.12	5.06	0.00	0.66	0.99	0.33	0.51

Model se zahrnutím pouze X_1

β_0	t_0	p	β_1	t_1	p	R^2
-6.92	-1.01	0.32	3.19	5.13	0.00	0.49

VÝSLEDKY LWS

Model se zahrnutím X_1, X_2

β_0	$t_0(w)$	p	β_1	$t_1(w)$	p	β_2	$t_2(w)$	p	R_W^2
0.92	0.14	0.89	2.06	4.91	0.00	-0.17	-0.38	0.71	0.94

Model se zahrnutím pouze X_1

β_0	t_0	p	β_1	t_1	p	R^2
-0.53	-0.13	0.89	2.01	5.56	0.00	0.94