The International Journal of
TRANSPORT & LOGISTICS
Medzinárodný časopis
DOPRAVA A LOGISTIKA

ISSN 1451-107X

LINEÁRNY REGRESNÝ MODEL ZÁVISLOSTI ŽIVOTNOSTI DOPRAVNÉHO PÁSU OD JEHO NIEKTORÝCH PARAMETROV

Miriam Andrejiová[1], Anna Pavlisková[2]

Kľúčové slová: životnosť dopravného pásu, regresný model, odhad parametrov, štatistická významnosť modelu.

Abstrakt:
Článok sa zaoberá klasickým lineárnym regresným modelom závislosti životnosti dopravného pásu od jeho niektorých parametrov: hrúbka krycej vrstvy, šírka a dĺžka pásu, rýchlosť prepravy a prepravované množstvo materiálu. Prvá časť článku sa venuje zostaveniu regresného modelu, bodovým a intervalovým odhadom parametrov. Druhá časť článku je venovaná overeniu štatistickej významnosti modelu a jednotlivých parametrov navrhovaného regresného modelu.

1. Úvod

Pod pojmom regresia rozumieme sledovanie vzájomných vzťahov medzi dvoma a viacerými premennými, pričom regresný model je matematický predpis, ktorý zjednodušene vyjadruje vzťah medzi premennými. Regresné modely môžeme deliť podľa rôznych hľadísk, napr. na lineárne a nelineárne regresné modely. V našom článku budeme uvažovať lineárny regresný model, ktorým chceme popísať závislosť životnosti dopravného pásu pomocou jeho niektorých parametrov, ktoré sme získali na základe prevádzkových záznamov z lomu Včeláre.

2. Regresný model a overenie modelu

Vzťah medzi vysvetľovanou premennou a k vysvetľujúcimi premennými môžeme vyjadriť lineárnym regresným modelom v tvare:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + b_4 x_{i4} + b_5 x_{i5} + \mathbf{K}\, b_k x_{ik} + e_i\,,$$

kde b_0 sa nazýva lokujúca konštanta a koeficienty b_j pre $j = 1, 2, \mathbf{K}, k$ sa nazývajú regresné koeficienty. Maticový zápis lineárneho regresného modelu je v tvare:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,,$$

kde $\mathbf{y}$ je n-prvkový stĺpcový vektor pozorovaných hodnôt vysvetľovanej premennej, $\mathbf{X}$ je matica typu $n \times (k+1)$ pozorovaných hodnôt vysvetľujúcich premenných, $\boldsymbol{\beta}$ je $(k+1)$prvkový stĺpcový vektor neznámych parametrov modelu a e je n-prvkový vektor náhodných chýb, ktoré predstavujú náhodnú zložku regresného modelu.

Bodovým odhadom lineárneho regresného modelu je regresná nadrovina

[1] **RNDr. Miriam Andrejiová, PhD.**, SjF TU v Košiciach, Katedra aplikovanej matematiky, Letná 9, 042 00 Košice, Slovensko, Tel.: +421 55 602 22 14, e-mail: miriam.andrejiova@tuke.sk
[2] **RNDr. Anna Pavlisková, PhD.**, SjF TU v Košiciach, Katedra aplikovanej matematiky, Letná 9, 042 00 Košice, Slovensko, Tel.: +421 55 602 22 08, e-mail: anna.pavliskova@tuke.sk

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + b_4 x_{i4} + b_5 x_{i5} + \ldots + b_k x_{ik},$$

kde $\hat{y}_i$ je vyrovnaná (teoretická) hodnota závisle premennej, b_0 je odhad lokujúcej konštanty a b_j je parciálny koeficient, ktorý je bodovým odhadom regresného koeficientu. Maticový zápis regresnej nadroviny je v tvare:

$$\hat{\mathbf{y}} = \mathbf{Xb},$$

kde $\hat{\mathbf{y}}$ je stĺpcový vektor vyrovnaných, teoretických hodnôt závisle premennej a $\mathbf{b}$ je stĺpcový vektor odhadnutých parametrov modelu.

Koeficienty parametrov $\boldsymbol{\beta}$ odhadneme pomocou metódy najmenších štvorcov, pre ktorú platí:

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 \rightarrow min,$$

kde $e_i = y_i - \hat{y}_i$ je rezíduum, pričom vektor rezíduí $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{Xb}$ je odhadom vektora náhodných chýb $\boldsymbol{\varepsilon}$. Prvky vektora $\mathbf{b}$ nájdeme riešením sústavy $(k+1)$ rovníc:

$$\begin{array}{ccccccccc}
\sum_{i=1}^{n} y_i & = & nb_0 & + & b_1 \sum_{i=1}^{n} x_{i1} & + & \cdots & + & b_k \sum_{i=1}^{n} x_{ik} \\
\sum_{i=1}^{n} x_{i1} y_i & = & b_0 \sum_{i=1}^{n} x_{i1} & + & b_1 \sum_{i=1}^{n} x_{i1}^2 & + & \cdots & + & b_k \sum_{i=1}^{n} x_{i1}.x_{ik} \\
\vdots & & \vdots & & \vdots & & & & \vdots \\
\sum_{i=1}^{n} x_{ik} y_i & = & b_0 \sum_{i=1}^{n} x_{ik} & + & b_1 \sum_{i=1}^{n} x_{i1}.x_{ik} & + & \cdots & + & b_k \sum_{i=1}^{n} x_{ik}^2
\end{array}$$

kde

$$X^T y = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik} y_i \end{pmatrix}, \quad X^T X = \begin{pmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \ldots & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}.x_{i2} & & \sum_{i=1}^{n} x_{i1}.x_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{i1}.x_{ik} & \sum_{i=1}^{n} x_{i2}.x_{ik} & & \sum_{i=1}^{n} x_{ik}^2 \end{pmatrix}.$$

Maticový tvar danej sústavy rovníc je $\mathbf{X^T y} = \mathbf{bX^T X}$, odkiaľ získame riešenie v tvare:

$$\mathbf{b} = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T y}.$$

V ďalšej časti sme skúmali závislosť *životnosti dopravných pásov* od niektorých jeho parametrov: *hrúbka krycej vrstvy*, *šírka* a *dĺžka pásu* a *množstvo prepravovaného materiálu* na 1m^2 (tabuľka 1). Lineárny regresný model, ktorý vystihuje závislosť životnosti pásu Z (závislá premenná) od nezávisle (vysvetľujúcich) premenných (hrúbka krycej vrstvu dopravného pásu H, šírka pásu S, dĺžka pásu D, rýchlosť R, prepravené množstvo PM) má tvar:

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + b_4 x_{i4} + b_5 x_{i5} + e_i, \text{ resp.}$$

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 + e \quad ,$$

kde $x_1 = H, x_2 = S, x_3 = D, x_4 = R, x_5 = PM$ a $\hat{y} = Z$.

Tab. 1 Parametre dopravných pásov

Č.	č. DP	H (mm)	S (m)	D (m)	R (m/s)	PM na t/hm^2	Z
1	301	8	1,2	90	1,6	4,315	11,3
2	302	8	1,2	100	1,6	3,600	10,7
3	304	7	1,2	99	1,4	2,180	10,7
4	306	9	1,2	7	1,4	25,714	40,4
5	307	9	1,2	7	1,5	15,429	40,4
6	309	7	1,0	32	1,6	3,750	20,7
7	310	9	1,0	42	1,6	4,286	18,0
8	312	7	1,0	27	1,8	9,444	22,5
9	322	12	1,2	14,4	1,4	12,500	28,1
10	381	7	0,8	15,8	1,4	20,222	32,9
11	382	7	1,2	26,2	1,4	1,096	20,9
12	383	6	1,0	24	1,4	1,400	23,9
13	401	6	0,8	126,6	1,4	1,007	11,6
14	402	6	0,8	202	1,4	1,485	9,2
15	403	7	0,8	107,6	1,4	2,091	12,6
16	404	6	0,8	196,6	1,4	2,517	9,3
17	405	7	1,2	160	1,7	0,428	8,4
18	507	7	1,4	78,6	1,4	0,863	14,2

Životnosť dopravného pásu bola vypočítaná podľa vzťahu $Z = \sqrt{\dfrac{2(c_z - c_v)}{c_r}}$, kde c_z je obstarávacia cena dopravného pásu, c_k je zostatková cena dopravného pásu a c_r sú rastúce náklady na údržbu dopravného pásu [2]. Bodový odhad lineárneho regresného modelu je:

$$\hat{y}_i = 22{,}5278 - 0{,}2522x_{i1} + 4{,}9979x_{i2} - 0{,}0673x_{i3} - 5{,}5522x_{i4} + 0{,}9470x_{i5}.$$

Lokujúca konštanta $b_0 = 22{,}5278$ predstavuje priemernú hodnotu životnosti dopravného pásu za predpokladu, že všetky vysvetľujúce premenné nadobúdajú nulovú hodnotu.

2.1. Overenie štatistickej významnosti modelu a parametrov modelu

V ďalšom kroku overíme, do akej miery lineárny regresný model odhadnutý metódou najmenších štvorcov vystihuje variabilitu závislej premennej a či vôbec vplyv niektorej vysvetľujúcej premennej na vysvetľovanú premennú (životnosť dopravného pásu) je relevantný. Použijeme **F – test štatistickej významnosti modelu**, ktorým overujeme hypotézy:

H_0 :*regresný model nie je štatisticky významný, (všetky regresné koeficienty sú nulové)*

proti

H_1 :*regresný model nie je štatisticky významný, (aspoň jeden regresný koeficient je nenulový).*

Za predpokladu platnosti nulovej hypotézy má testovacia charakteristika

$$F = \frac{MS_M}{MS_R} = \frac{\dfrac{SS_M}{k}}{\dfrac{SS_R}{n-k-1}} = \frac{(n-k-1)\cdot\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{k\cdot\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Fischerovo F – rozdelenie s k a $(n-k-1)$ stupňami voľnosti ($k=5$). Na hladine významnosti a zamietame nulovú hypotézu, ak $F > F_{1-a}(k;n-k-1)$. V prípade zamietnutia nulovej hypotézy aspoň jedna vysvetľujúca premenná má štatisticky významný vplyv na testovanú vysvetľovanú premennú.

Súčet štvorcov SS_M predstavuje variabilitu vyrovnaných hodnôt okolo priemeru $\bar{y}$, SS_R predstavuje rozdiel skutočnej a vyrovnanej hodnoty (rezíduum). Celková variabilita SS_T závislej premennej sa môže vyjadriť v tvare $SS_T = SS_M + SS_R$. V Tab. 2 je výsledná tabuľka analýza rozptyly pre navrhovaný regresný model.

***Tab. 2** Tabuľka analýzy rozptylu*

Variabilita	Súčet štvorcov	Stupne voľnosti	Priemerný štvorec	Testovacia charakteristika *F*
Vysvetlená modelom	$SS_M = 1598{,}769$	$df_M = 5$	$MS_M = 319{,}7538$	$F = \frac{MS_M}{MS_R}$
Nevysvetlená modelom	$SS_R = 254{,}4489$	$df_R = 18-5-1 = 12$	$MS_R = 21{,}2041$	
Celková	$SS_T = 1853{,}218$	$df_T = 17$		

Hodna testovacej charakteristiky je $F = \frac{MS_M}{MS_R} = 15{,}0798$, kritická hodnota $F_{0,95}(5;12) = 3{,}106$.

Pretože $F = 15{,}08 > F_{0,95}(5;12) = 3{,}106$ ($p = 8{,}14.10^{-6} < a$), nulovú hypotézu zamietame na hladine významnosti $a = 0{,}05$ a môžeme predpokladať, že navrhovaný regresný model je štatisticky významný; teda aspoň jedna z vysvetľujúcich premenných významne ovplyvňuje životnosť dopravných pásov.

Štatistickú významnosť jednotlivých parametrov overíme pomocou **testu štatistickej významnosti regresného koeficientu** b_j. Testujeme H_0 *:regresný koeficient nie je štatisticky významný, resp. (* $b_j = 0$ *)* proti H_1 *: regresný koeficient je štatisticky významný, resp. (* $b_j \neq 0$ *).*

Testovacia charakteristika je $t = \frac{b_j}{s_{b_j}}$, kde b_j je bodový odhad koeficientu a s_{b_j} je ich smerodajná odchýlka. Nulovú hypotézu zamietame na hladine významnosti a, ak $|t| > t_{1-\frac{a}{2}}(n-k-1)$.

V Tab. 3 sú odhady parametrov, intervalový odhad parametrov a vyhodnotenie prínosu vysvetľujúcich premenných v navrhovanom modeli.

***Tab. 3** Odhady parametrov*

Parameter	**Bodový odhad**	**s**	**t**	**p-hodnota**	**95% interval**	
					DH	**HH**
b_0	22,5278	14,3037	1,575	0,1412	-8,6372	53,6929
b_1 (H)	-0,2522	1,0151	-0,248	0,8080	-2,4638	1,9594
b_2 (S)	4,9979	6,8539	0,729	0,4799	-9,9356	19,9313
b_3 (D)	-0,0673	0,0238	-2,825	0,0153 *	-0,1192	-0,0154
b_4 (R)	-5,5522	8,9268	-0,619	0,5477	-24,9724	13,9274
b_5 (PM)	0,9470	0,2011	4,710	0,00045 *	0,5089	1,3851

Intervalový odhad parametrov modelu môžeme použiť aj k štatistickému testovaniu významnosti parametrov modelu. V prípade, že nula leží v danom intervale spoľahlivosti parametra, potom daný parameter je štatisticky nevýznamný (v našom prípade sa takmer všetky koeficienty, okrem b_3 a b_5, javia ako štatisticky nevýznamné).

Bodovým odhadom viacnásobného koeficientom korelácie je výberový viacnásobný koeficient korelácie $r^2 = 0{,}8627$. Neskresleným odhadom viacnásobného koeficientu determinácie, ktorý zohľadňuje počet vysvetľujúcich parametrov modelu, sa nazýva upravený (korigovaný) koeficient determinácie $r^2_{adj} = 0{,}8055$.

3. Záver

Porovnaním hodnôt vyplýva, že nulovú hypotézu zamietame v prípade vysvetľujúcej premennej *dĺžka dopravného pásu* D a *prepravovaného množstva* PM (ozn.*, $p < a$). V ostatných prípadoch nulovú hypotézu nezamietame a vysvetľujúce premenné *šírka pásu*, *hrúbka krycej vrstvy* a *rýchlosť dopravného pásu* môžeme z modelu vylúčiť. Daným regresným modelom vieme vysvetliť variabilitu premennej *životnosť* Z na 86,27%. Zvyšných 13,73% spôsobujú činitele nezahrnuté do modelu alebo iné vysvetľujúce premenné a náhodné vplyvy. V ďalšej časti odporúčame skúmať splnenie podmienok pre náhodnú zložku a diagnostikovanie extrémnych a vplyvných hodnôt modelu.

Príspevok bol spracovaný s podporou grantových projektov VEGA č.1/0543/10 a VEGA č.1/0864/10.

Literatúra:

[1] Pavlisková, A., Jadroňová, M.: Modely obnovy v pásovej doprave. DOPRAVA A LOGISTIKA, mimoriadne číslo, 2006, ISSN 1451-107X.
[2] Pavlisková, A., Jadroňová, M.: Využitie mat. modelu prevádzkových nákladov na výpočet životnosti dop. pásov. Výrobné inžinierstvo, č.4, roč. V., str. 61-62,72, ISSN 1335-7972.
[3] Šoltés, E.: Regresná a korelačná analýza s aplikáciami. Bratislava, 2007. Iura Edition. 287 s. ISBN 978-80-8078-163-7.

Recenzia/Review: *prof. Ing. Daniela Marasová, CSc.*