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

ISSN 1451-107X

ANALÝZA NÁHODNEJ ZLOŽKY LINEÁRNEHO REGRESNÉHO MODELU ZÁVISLOTI ŽIVOTNOSTI DOPRAVNÉHO PÁSU OD JEHO NIEKTORÝCH PARAMETROV

Miriam Andrejiová[1], Anna Pavlisková[2]

Kľúčové slová: regresný model, náhodná zložka modelu, studentizované rezíduá, Durbin – Watsonova charakteristika, koeficient autokorelácie 1. stupňa.

Abstrakt:
Článok sa zaoberá analýzou náhodnej zložky klasického lineárneho regresného modelu závislosti životnosti dopravného pásu od niektorých jeho parametrov, a to predovšetkým overením podmienok nezávislosti, normality a homoskedasticity rezíduí.

1. Úvod

V článku „Lineárny regresný model závislosti životnosti dopravného pásy od jeho niektorých parametrov“ sme skúmali závislosť *životnosti dopravných pásov* (Z) od niektorých jeho parametrov: *hrúbka krycej vrstvy (H)*, *šírka pásu (S), dĺžka pásu* (D), r*ýchlosť* (R) *a množstvo prepravovaného materiálu* na 1m^2 (PM).

Vytvorený model funkčnej závislosti má tvar

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

kde b_0 sa nazýva lokujúca konštanta, b_j pre $j = 1, 2, \ldots, 5$ sa nazývajú regresné koeficienty a e_i náhodné chyby. Bodový odhad regresného modelu môžeme zapísať aj v tvare

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 \text{, kde } x_1 = H, x_2 = S, x_3 = D, x_4 = R, x_5 = PM \text{ a } \hat{y} = Z \,.$$

Zistili sme, že závislosť životnosti dopravného pásy od nami vybraných parametrov (Tab.1) sa dá popísať bodovým odhadom v tvare

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

V regresnom modeli sa predpokladá, že náhodné chyby e_i sú navzájom nezávislé, s normálnym rozdelením so strednou hodnotou nula a konštantným rozptylom s_e^2. Vlastnosti náhodných chýb sú:

- $E(e_i) = 0$ (*stredná hodnota náhodných chýb* e_i *je rovná nula*),
- $D(e_i) = s_e^2$ (*rozptyl náhodnej zložky* e_i *je konštantný, homoskedasticita*)
- $cov(e_i; e_j) = 0$ pre všetky $i \neq j$ (*vzájomná lineárna nezávislosť ľubovoľnej dvojice náhodných chýb*)

---

[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

§ $e_i = N(0; s_e^2)$, čo znamená, že *náhodné zložky* e_i *majú normálne rozdelenie*.

Ak je navrhovaný regresný model vhodný, potom rezíduá e_i by mali spĺňať vlastnosti náhodných chýb e_i.

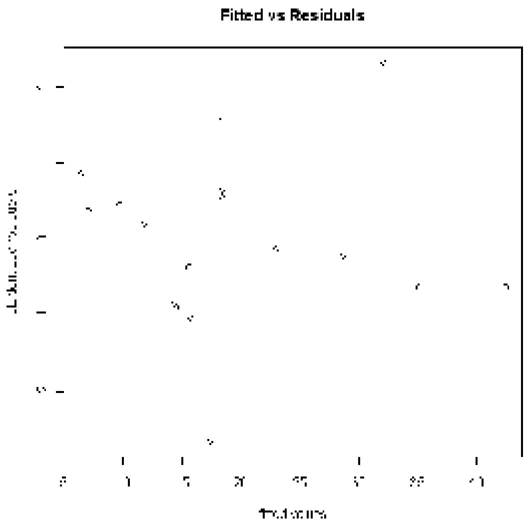
2. Analýza náhodnej zložky regresného modelu

2. 1 Grafická analýza rezíduí

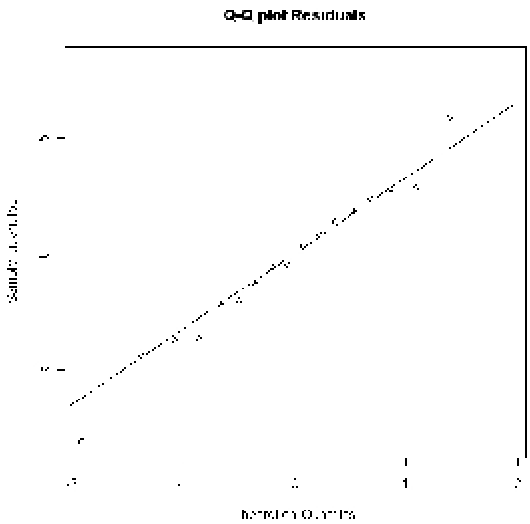
Grafickú analýzu rezíduí môžeme vykonať pomocou grafov rezíduí, najčastejšie studentizovaných rezíduí e_{Si}, vzhľadom na hodnoty regresnej funkcie $\hat{y}_i$, teda graf bodov $[\hat{y}_i; e_{Si}]$.

Ak rezíduá spadajú do horizontálneho pásu sústredeného okolo nuly a v intervale $(-2; 2)$ sa nachádzajú takmer všetky studentizované rezíduá a mimo intervalu $(-3; 3)$ sa objavujú rezíduá ojedinele (menej ako 1%), potom takéto rozmiestnenie rezíduí nenaznačuje porušenie predpokladov kladených na náhodnú zložku.

Z grafu (Obr. 1) vyplýva, že studentizované rezíduá sa sústreďujú okolo 0 a ich početnosti sa so zväčšujúcou vzdialenosťou od hodnoty 0 zmenšujú. V intervale $(-2; 2)$ sa nachádza viac ako 95% hodnôt a preto graf nenaznačuje porušenie predpokladov o náhodných chybách.



Obr. 1 *Rezídua vs hodnoty*



Obr. 2 *O –Q graf*

2.2 Overenie normality

Testujeme H_0 : *rezíduá sú z normálneho rozdelenia*

proti H_1 : *rezíduá nie sú z normálneho rozdelenia.*

Na overenie normality využijeme najprv grafickú analýzu pomocou tzv. diagnostických grafoch, napr. pomocou Q-Q grafu (Obr. 2).

Ak grafická analýza nevedie k jednoznačnému rozhodnutiu, testujeme normalitu pomocou štatistických testov: napr. Shapiro – Wilkov test, Pearsonov test, Cramer von Misesov test a iné.

Testovacia charakteristika *Shapiro – Wilkovho testu* W sa rovná 0,9664, *p* - hodnota pre obojstranný test je 0,7278. Pretože $p \geq a$, nulovú hypotézu H_0 nezamietame na hladine významnosti $a = 0{,}05$ (resp. $a = 0{,}01$) a môžeme predpokladať, že náhodné chyby majú normálne rozdelenie.

Rovnaký výsledok získame aj pomocou Cramérov-von Misesov test ($p = 0{,}9316$) a Andersonovho-Darlingovho testu normality ($p = 0{,}9046$).

2.3 Overenie homoskedasticity

Testujeme H_0 : *homoskedasticita rezíduí* proti H_1 : *heteroskedasticita rezíduí.*

Na overenie predpokladu o konštantnom rozptyle náhodných chýb použijeme *Goldfeldov – Quandtov tes* [5]. Súbor vyrovnaných hodnôt sme usporiadali zostupne a rozdelili na 2 časti, v ktorých sme vypočítali súčet štvorcov studentizovaných rezíduí.

Pre testovaciu charakteristiku platí $F = \frac{\sum_{i \in n_2} e_{Si}^2}{\sum_{i \in n_1} e_{Si}^2}$, kde volíme $\sum_{i \in n_2} e_{Si}^2 > \sum_{i \in n_1} e_{Si}^2$. Hodnota testovacej charakteristiky je $F = 1{,}1503$. Nulovú hypotézu o homoskedasticite zamietame, ak $F > F_{1-a}(n_2 - k - 1; n_1 - k - 1)$.

Pretože $F = 1{,}1503 < F_{0{,}95}(9; 9) = 3{,}179$, na hladine významnosti $a = 0{,}05$ nezamietame nulovú hypotézu o konštantnom rozptyle náhodných chýb. V prípade hladiny významnosti $a = 0{,}01$ platí.

Homoskedasticitu rezíduí sme overili aj pomocou *Breuschovho-Paganovho testu*, podľa ktorého rovnako ako v predchádzajúcom teste, nemôžeme zamietnuť nulovú hypotézu o konštantnom rozptyle náhodných chýb ($p = 0{,}5286$).

Tab. 1 *Parametre dopravných pásov a rezíduá*

Č.	č. DP	H (mm)	S (m)	D (m)	R (m/s)	PM na t/hm^2	Z	Zt*	e_i	e_{Si}
1	301	8	1,2	90	1,6	4,315	11,3	15,703	-4,403	-1,033
2	302	8	1,2	100	1,6	3,600	10,7	14,353	-3,653	-0,852
3	304	7	1,2	99	1,4	2,180	10,7	14,432	-3,732	-0,889
4	306	9	1,2	7	1,4	25,714	40,4	42,404	-2,004	-0,634
5	307	9	1,2	7	1,5	15,429	40,4	32,112	8,288	2,321
6	309	7	1,0	32	1,6	3,750	20,7	18,323	2,377	0,552
7	310	9	1,0	42	1,6	4,286	18,0	17,653	0,237	0,086
8	312	7	1,0	27	1,8	9,444	22,5	22,947	-0,447	-0,135
9	322	12	1,2	14,4	1,4	12,500	28,1	28,636	-0,536	-0,211
10	381	7	0,8	15,8	1,4	20,222	32,9	35,117	-2,217	-0,629
11	382	7	1,2	26,2	1,4	1,096	20,9	18,304	2,596	0,641
12	383	6	1,0	24	1,4	1,400	23,9	17,992	5,908	1,582
13	401	6	0,8	126,6	1,4	1,007	11,6	9,718	1,882	0,449
14	402	6	0,8	202	1,4	1,485	9,2	17,329	-8,129	-2,653
15	403	7	0,8	107,6	1,4	2,091	12,6	11,770	0,830	0,197
16	404	6	0,8	196,6	1,4	2,517	9,3	6,438	2,862	0,870
17	405	7	1,2	160	1,7	0,428	8,4	7,012	1,388	0,374
18	507	7	1,4	78,6	1,4	0,863	14,2	15,557	-1,357	-0,372

Poznámka: Z* sú vyrovnané hodnoty životnosti dopravných pásov, ktoré sú získané regresným modelom.

2.4 Overenie nezávislosti náhodných chýb

Nezávislosť náhodných chýb sme overili pomocou koeficienta autokorelácie 1. stupňa r_1 a pomocou najpoužívanejšieho testu autokorelácie – Durbin – Watsonovej charakteristiky. Vychádzali sme z rezíduí usporiadaných podľa veľkosti regresnej funkcie, pričom použili sme aj klasické e_i aj studentizované e_{Si} rezíduá.

Bodovým odhadom *koeficienta autokorelácie 1. stupňa* je výberový koeficient autokorelácie, pre ktorý platí

$$r_1 = \frac{\sum_{i=2}^{n} e_i \cdot e_{i-1}}{\sum_{i=1}^{n} e_i^2} = 0{,}1789 \text{ (klasické rezíduá), resp. } r_{1S} = \frac{\sum_{i=2}^{n} e_{Si} \cdot e_{S(i-1)}}{\sum_{i=1}^{n} e_{Si}^2} = 0{,}1067 \text{ (studentizované rezíduá).}$$

Prostredníctvom testu štatistickej významnosti autokorelácie 1. stupňa otestujeme nezávislosť náhodných chýb. Testujeme H_0: $r_1 = 0$ oproti $H_1: r_1 \neq 0$ (resp. H_0*: koeficient autokorelácie nie je* *štatisticky významný; chyby sú navzájom nezávislé* oproti H_1: *koeficient autokorelácie je štatisticky významný; chyby sú navzájom závislé, autokorelované*).

Nulovú hypotézu zamietame na hladine významnosti a, ak $|r_1| \geq |r_a(n)|$, kde $r_a(n)$ je tabelovaná kritická hodnota. Pretože $|0{,}1789| < |r_{0,05}(18)| = 0{,}299$ (resp. $|0{,}1067| < |r_{0,05}(18)| = 0{,}299$). V prípade hladiny významnosti $a = 0{,}01$ je $r_{0,01}(18) = 0{,}432$. Z výsledku testu vyplýva, že koeficient autokorelácie 1. stupňa nie je štatisticky významný a predpoklad o nezávislosti chýb môžeme považovať za splnený.

Testovacia Durbin-Watsonova charakteristika je daná vzťahom

$$D-W = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}, \text{ resp. } D-W = \frac{\sum_{i=2}^{n}\left(e_{Si} - e_{S(i-1)}\right)^2}{\sum_{i=1}^{n} e_{Si}^2}.$$

V našom prípade pre klasické rezíduá je $D-W = 1{,}559$, resp. pre studentizované rezíduá $D-W = 1{,}7275$. Vo všeobecnosti sú hodnoty $D-W$ z intervalu $\langle 0, 4\rangle$. V praxi môžeme postupovať podľa zjednodušeného pravidla, podľa ktorého ak hodnota testovacej charakteristiky $D-W$ leží v intervale (1,4; 2,6), rezíduá nevykazujú autokoreláciu [1]. Ak nevieme rozhodnúť, odporúča sa použiť kritické hodnoty [5].

3. Záver

Z výsledkov analýzy náhodných chýb navrhnutého regresného modelu závislosti životnosti dopravného pásu od jeho niektorých parametrov vyplýva, že je splnený predpoklad normality náhodných chýb a homoskedasticita. Podľa Durbin – Watsonovej charakteristiky a koeficienta autkorelácie je potvrdená aj nezávislosť náhodných chýb. V ďalšej analýze regresného modelu by bolo zaujímavé venovať pozornosť diagnostikovaniu extrémnych a vplyvných hodnôt, ktoré môžu výrazne vplývať na odhad regresného modelu a jeho parametrov.

Príspevok bol spracovaný s podporou grantového projektu VEGA č.1/0543/10 a VEGA č.1/0864/10.

Literatúra:

[1] Hudec, O. – Sisáková, J. – Rataľová, A. – Želinský, T.: Štatistické metódy v ekonomických vedách. Košice, 2007. Elfa, s.r.o.,2007. 196 s. ISBN 978-80-8086-059-2

[2] Meloun, M. – Militký, J.: Statistická analýza experimentálních dat. Praha, 2004. ACADEMIA, 953 s. ISBN 80-200-1254-0

[3] Pavlisková, A., Jadroňová, M.: Modely obnovy v pásovej doprave. DOPRAVA A LOGISTIKA, mimoriadne číslo, 2006, ISSN 1451-107X.

[4] 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.

[5] Šoltés, E.: Regresná a korelačná analýza s aplikáciami. Bratislava, 2007. Iura Edition. 287 s. ISBN 978-80-8078-163-7.

Recenzia/Review: *doc. Ing. Vieroslav Molnár, PhD.*