



EMPIRICKÝ PRÍSTUP A JEHO POSTAVENIE V EXPERIMENTÁLNOHOM MODELOVANÍ

AN EMPIRICAL APPROACH AND ITS POSITION IN EXPERIMENTAL MODELING

Peter TREBUŇA

Abstract

An empirical approach is based on analysis and evaluation of generalizations of positive and negative experiences of practice. In literature it is usually presented by the representatives of the major points of view empiricist relating to the management of each site. Most representatives of the empirical approach seek to draw on the study of management practices specific recommendations for management action.

Key words

Empirical Model, Analysis of Variance (ANOVA), Distribution.

Úvod

Empirický prístup k zostrojeniu modelu je nevyhnutný vtedy ak už poznáme dominantné premenné ovplyvňujúce proces, ale nepoznáme model procesu.

Nájdanie empirického modelu predpokladá [1,6]:

- výber vhodného modelu,
- vykonanie experimentu k získaniu potrebných údajov a informácií,
- aproximácia modelu, t. j. vyčíslenie koeficientov modelu,
- overenie výsledku.

Model môže byť lineárny alebo nelineárny vzhľadom ku koeficientom modelu. Všeobecný tvar lineárneho modelu s nezávislou premennou je:

$$\eta = \beta_0 \cdot f_0(x) + \beta_1 \cdot f_1(x) + \dots + \beta_n \cdot f_n(x),$$

kde β_i sú koeficienty modelu a $f_i(x)$ sú známe funkcie pričom $i = 1, 2, \dots, n$.

Lineárny model

Linearizáciou modelu rozumieme postup, ktorým sa nelineárny model matematickými úpravami pretransformuje na lineárny model.

Linearizovaný model nie je ekvivalentný s pôvodným nelineárnym modelom ak sa transformuje nelineárnou transformáciou závisle premennej. Ak údaje (x_i, y_i) popisujú proces, potom môžeme určiť tvar závislosti $y = f(x)$, pre ktorý táto závislosť nie je závislosťou funkčnou (ako v matematike) s pravdepodobnosťou $p=1$ ale funkciou voľnejšou, kde $0 < p < 1$.

Neexistuje univerzálny spôsob nájdania vhodného typu regresnej funkcie ale vo väčšine prípadov sa využívajú:

- skúsenosti a znalosti skúmaného procesu,
- vynesenie bodov do grafov a výber podľa katalógu kriviek,
- iných pomocných kritérií.

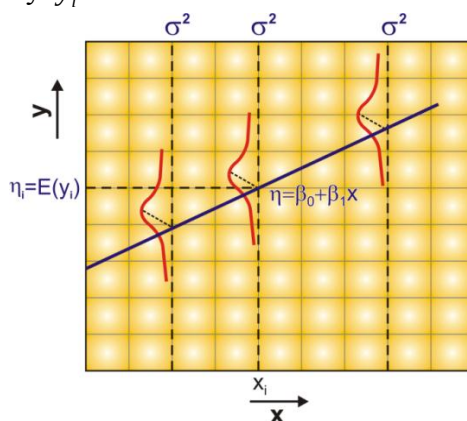


Lineárny model tvaru $\eta = \beta_0 + \beta_1 \cdot x$

sa vyznačuje tým, že derivácia $\frac{d\eta}{dx}$ je konštantná, čo sa využíva k nájdeniu vhodného tvaru regresnej funkcie.

Predpokladajme, že v modeli $\eta = \beta_0 + \beta_1 x$:

- pre každú pevnú hodnotu x_i majú hodnoty y_i normálne rozdelenie (pozri obr.1),
- každé z týchto rozdelení má tiež rozptyl σ^2 , tzn. $D(y_i | x_i) = \sigma^2$,
- priemerné hodnoty týchto rozdelení spĺňajú vzťah $E(\bar{y}_i | x_i) = \eta_i = \beta_0 + \beta_1 x_i$,
- rôzne merané veličiny y_i sú na sebe nezávislé.



Obr. 1 Konštantný rozptyl meraných veličín

Model možno zapísať v tvare:

$$\bar{y}_i = \beta_0 + \beta_1 x_i + e_i,$$

kde e_i je tzv. náhodná chyba, pri ktorej platí: $E(e_i) = 0, D(e_i) = E(e_i^2) = \sigma^2, E(e_i e_j) = 0$ pre $i \neq j$. Koeficienty β_0, β_1 nepoznáme ale môžeme ich odhadnúť metódou najmenších štvorcov, teda z podmienky minimalizácie výrazu:

$$Q = \sum_{i=1}^n (\bar{y}_i - \eta_i)^2 p_i = \sum_{i=1}^n p_i (\bar{y}_i - \beta_0 - \beta_1 x_i)^2,$$

kde p_i je počet opakovaných meraní na úrovni x_i a n je celkový počet úrovní meraní [2].

Celkový počet meraní je však $\sum_{i=1}^n p_i$.

Odhady všetkých veličín budeme označovať strieškou, tzn. $\beta_0 = b_0, \beta_1 = b_1$.

Na nasledujúcom obr.2 sú pre lepšiu predstavu znázornené dve priamky, a to graf modelu ($\eta = \beta_0 + \beta_1 x$) a graf toho, čo skutočne vypočítame ($y = b_0 + b_1 x$). Navyiac sú tam tieto body:

$[x_i, y_{ij}]$ - jedno z meraní na úrovni x_{ij} , $1 \leq j \leq p_i$

$[x_i, \bar{y}_i]$ - priemerná hodnota y na úrovni x_i

$[x_i, \eta_i]$ - odhad hodnoty $\bar{y}_i$ na úrovni x_i

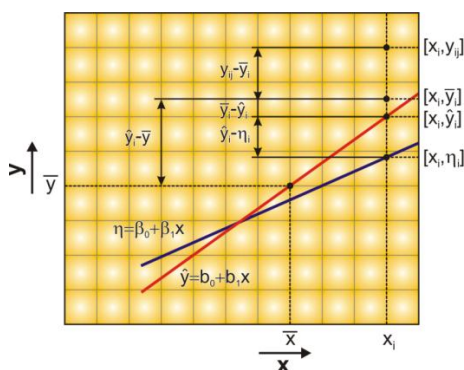


$[x_i, \eta_i]$ - stredná hodnota hodnoty $\bar{y}_i$ pre $x = x_i$

Lineárny model $\eta = \beta_0 + \beta_1 x$ sa väčšinou zapisuje v tvare $\eta = \beta_0 + \beta_1 (x - \bar{x})$, nakoľko odhady b_0, b_1 koeficientov β_0, β_1 možno získať bez nutnosti riešenia sústavy rovníc. Aby výraz Q mal minimálnu hodnotu, musí platiť $\frac{\partial Q}{\partial \beta_0} = 0$, $\frac{\partial Q}{\partial \beta_1} = 0$, teda

$$-2 \sum_{i=1}^n p_i [\bar{y}_i - \beta_0 - \beta_1 (x_i - \bar{x})] = 0,$$

$$-2 \sum_{i=1}^n p_i [\bar{y}_i - \beta_0 - \beta_1 (x_i - \bar{x})] (x_i - \bar{x}) = 0.$$



Obr.2 Graf modelov dvoch priamok

Ak zameníme koeficientom β_0, β_1 ich odhady, potom po úprave dostaneme tzv. normálne rovnice (sú vynechané sčítacie indexy pre $i = 1$ až n) :

$$\sum p_i \bar{y}_i = b_0 \sum p_i + b_1 \sum p_i (x_i - \bar{x})$$

$$\sum p_i \bar{y}_i (x_i - \bar{x}) = b_0 \sum p_i (x_i - \bar{x}) + b_1 \sum p_i (x_i - \bar{x})^2.$$

Keďže pre $\bar{x} = \frac{\sum p_i x_i}{\sum p_i}$, je

$$\sum p_i (x_i - \bar{x}) = \sum p_i x_i - \bar{x} \sum p_i = \sum p_i x_i - \frac{1}{\sum p_i} \sum p_i x_i \sum p_i = \sum p_i x_i - \sum p_i x_i = 0,$$

z normálnych rovníc je možné priamo vypočítať:

$$\beta_0 = b_0 = \frac{\sum_{i=1}^n p_i \bar{y}_i}{\sum_{i=1}^n p_i} = \bar{y}, \quad \beta_1 = b_1 = \frac{\sum_{i=1}^n p_i (x_i - \bar{x})}{\sum_{i=1}^n p_i (x_i - \bar{x})^2}$$

Analýza rozptylu, označovaná taktiež ako ANOVA (Analysis of variance), sa vo všeobecnosti používa buď ako samostatná technika alebo ako postup umožňujúci analýzu zdrojov variability v lineárnych statických modeloch. V ďalšom budeme analyzovať rozptyl, aby sme získali informácie o zdrojoch variability empirického modelu. Pri analýze rozptylu postupujeme tak, že celkovú variabilitu výsledkov pokusov rozdelíme na samostatné zložky, ktoré prislúchajú skúmaným činiteľom. Všeobecne rozdelíme príčiny variability na dve skupiny:



- na príčiny, ktoré pôsobia systematicky, t.j. zmeny príčiny podmieňujú relatívne pravidelné zmeny hodnôt skúmaného činiteľa (znaku). Nazývame ich faktory variability,
- na príčiny, ktoré vyvolávajú zmeny skúmaného znaku nepravidelne, náhodne. Nazývame ich náhodné vplyvy [3].

Súčet štvorcov na ľavej strane (tzv. úplný súčet štvorcov) udáva mieru rozdielov medzi experimentálnymi hodnotami a strednými hodnotami premennej y pre dané hodnoty x_i . Prvý člen sprava je súčet štvorcov odchýlok vo vnútri jednotlivých úrovní $x = x_i$ (tzv. súčet štvorcov chýb). Druhý člen predstavuje súčet štvorcov odchýlok od empirickej krivky (reziduálny súčet štvorcov). Tretí, resp. štvrtý člen, je súčet štvorcov odchýlok b_0 od β_0 , resp. b_1 od β_1 .

Súčet štvorcov chýb sa dá interpretovať ako miera experimentálnych chýb, ktorých sme sa dopustili postupne pri jednotlivých meraniach na úrovni x_i . Reziduálny súčet štvorcov je mierou efektívnosti (adekvátnosti) lineárneho modelu pre aproximáciu nameraných hodnôt.

Pretože

$$E\left(\frac{1}{n-2} \sum_i p_i (\bar{y}_i - \hat{y}_i)^2\right) = \frac{1}{n-2} E(\sigma^2 \chi^2) = \frac{\sigma^2}{n-2} E(\chi^2) = \sigma^2,$$

$$\text{je možné veličinu } s_r^2 = \frac{1}{n-2} \sum_{i=1}^n p_i (\bar{y}_i - \hat{y}_i)^2$$

považovať za odhad rozptylu σ^2 , tzn. $\hat{\sigma}^2 = s_r^2$.

$$\text{Podobne veličina } s_e^2 = \frac{1}{\sum_{i=1}^n p_i - n} \sum_{i=1}^n \sum_{j=1}^{p_i} (y_{ij} - \bar{y}_i)^2 \text{ je odhad rozptylu } \sigma^2.$$

Tab.1 Analýza rozptylu vzhľadom k lineárnemu modelu

zdroj variability	odchýlka b_0 od β_0	odchýlka b_1 od β_1	odchýlka od priamky - nelinearita	chyba merania	celkový odhad rozptylu
súčet štvorcov	$(b_0 - \beta_0)^2 \cdot \sum_i p_i$	$(b_1 - \beta_1)^2 \cdot \sum_i p_i \cdot (x_i - \bar{x})^2$	$\sum_i p_i \cdot (\bar{y}_i - \hat{y}_i)^2$	$\sum_i \sum_j (y_{ij} - \bar{y}_i)^2$	$\sum_i \sum_j (y_{ij} - \eta_i)^2$
Stupne voľnosti	1	1	$n-2$	$\sum_i p_i - n$	$\sum_i p_i$
priemerný štvorec	$E(b_0) \cdot \sum_i p_i$	$E(b_1) \cdot \sum_i p_i \cdot (x_i - \bar{x})^2$	$s_r^2 = \frac{\sum_i p_i (\bar{y}_i - \hat{y}_i)^2}{n-2}$	$s_e^2 = \frac{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2}{\sum_i p_i - n}$	
testovacia veličina F			$F = \frac{s_r^2}{s_e^2}$		
kvantil F-rozdelenia			$F\left(1-\alpha, n-2, \sum_i p_i - n\right)$		
H_0			$\eta = \beta_0 + \beta_1 (x - \bar{x})$		



Bodové odhady majú malú informačnú hodnotu, dokiaľ sa neuvedie interval spoľahlivosti (kofidenčný interval) pre tieto náhodné veličiny [5].

V ďalšom určíme nepresnosti jednotlivých výpočtových hodnôt závisle premennej s istou spoľahlivosťou, ale bez vzájomného vzťahu k ďalším výpočtovým hodnotám. Ak vynesieme do grafu tieto intervaly pre všetky x , potom krajné body týchto intervalov vytvoria pás spoľahlivosti. Tento pás nás neinformuje o nepresnosti pre celú regresnú priamku ako celku, ale len o jednotlivých intervaloch, vždy pre jedinú hodnotu nezávisle premennej x .

Pás spoľahlivosti pre celú regresnú priamku, tzn. pre všetky body súčasne, možno určiť a následne testovať tzv. nulovou hypotézou pre t-rozdelenie:

$$H_0 : \beta_0 = b_0^*, \text{ resp. } \beta_1 = b_1^*, \text{ proti } H_1 : \beta_0 \neq b_0^*, \text{ resp. } \beta_1 \neq b_1^*,$$

kde b_0^*, b_1^* sú dané čísla. Testovanou veličinou je $t = \frac{|b_0 - b_0^*|}{\sqrt{\hat{D}(b_0)}}$, respektíve $t = \frac{|b_1 - b_1^*|}{\sqrt{\hat{D}(b_1)}}$, kde

$$b_0 = \hat{\beta}_0, \quad b_1 = \hat{\beta}_1.$$

Ak $t \geq t\left(1 - \frac{\alpha}{2}, r\right)$, kde $t(\cdot)$ je $1 - \alpha/2$ kvantil Studentovho rozdelenia s počtom stupňov voľnosti $r = \sum_i p_i - 2$, potom so spoľahlivosťou $100(1 - \alpha/2)\%$ zamietame H_0 .

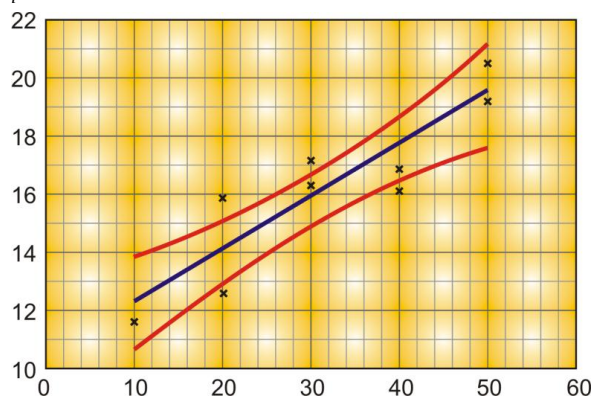
Pre $b_0^* = 0$, resp. $b_1^* = 0$, je tento test zhodný s testom v tabuľke analýzy rozptylu.

Bodový odhad lineárneho modelu je $\hat{y} = \hat{b}_0 + \hat{b}_1 x$. Intervalový odhad vypočítanej hodnoty z tohto modelu pre pevné x na hladine $1 - \alpha$ je preto:

$$\hat{b}_0 + \hat{b}_1 x \pm t\left(1 - \frac{\alpha}{2}, r\right) \cdot s.d(x),$$

$$r = \sum_i p_i - 2, t(\cdot) \text{ je príslušný kvantil t-rozdelenia, } s \text{ je odhad smerodajnej odchýlky } \sigma \text{ a}$$

$$d(x) = \sqrt{\frac{1}{\sum_i p_i} + \frac{(x - \bar{x})^2}{\sum_i p_i (x_i - \bar{x})^2}}.$$



Obr.3 Pás spoľahlivosti okolo regresnej priamky pre jednu hodnotu x

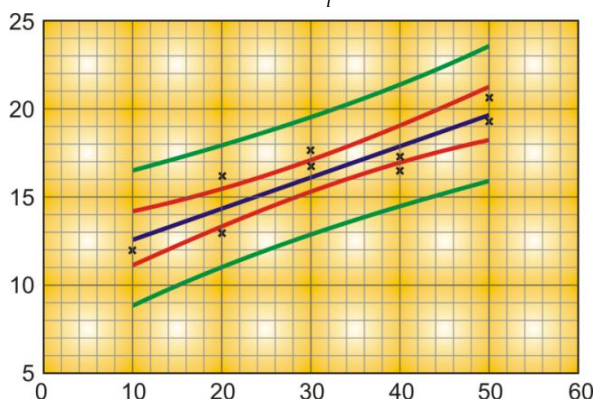
Tieto intervalové odhady vytvoria tzv. pás spoľahlivosti okolo regresnej priamky. Je to pás, ktorého šírka pre pevne zvolenú hodnotu x udáva s istou spoľahlivosťou nepresnosť vypočítanej hodnoty závisle premennej zo zvoleného modelu.



Pás spoľahlivosti, určuje nepresnosť regresnej priamky nie pre všetky x súčasne, ale vždy pre jediné, i keď ľubovoľné x . Pás spoľahlivosti pre všetky x súčasne na hladine $1-\alpha$ je daný vzťahom

$$b'_0 + b_1 x \pm s \cdot \sqrt{2 \cdot F(1-\alpha, 2, r)} \cdot d(x),$$

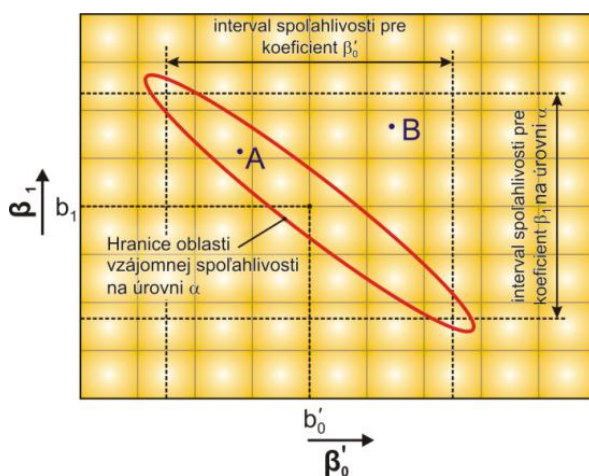
kde $d(x)$ sa vypočíta z predchádzajúceho vzťahu pre výpočet $d(x)$ a $F(\cdot)$ je $(1-\alpha)$ -kvantil F-rozdelenia s počtom stupňov voľnosti 2 a $r = \sum_i p_i - 2$.



Obr.4 Pás spoľahlivosti pre všetky x súčasne, ale vždy pre jedno

Ak chceme zistiť, či namerané hodnoty popisuje lineárny model, v ktorom sa súčasne môžu meniť odhady koeficientov β_0 a β_1 (tzn. priamka sa súčasne posúva a otáča), potom je potrebné vykonať odhad oblasti spoľahlivosti pre β_0 a β_1 . Obdĺžnik vytvorený odhadmi intervalov spoľahlivosti a odhad oblasti spoľahlivosti môžu obsahovať rozdielne hodnoty odhadov koeficientov β_0 a β_1 .

Z obr. 5 je zrejmé, ako môže vyzerat' oblasť spoľahlivosti pre dva regresné koeficienty.



Obr.5 Oblasť spoľahlivosti pre dva regresné koeficienty



Oba body A a B sa nachádzajú vo vnútri obdĺžnika vytvoreného zo samotných intervalov spoľahlivosti. Bod B však leží mimo hranice oblasti vzájomnej spoľahlivosti, tzn. príslušné odhady $\hat{\beta}_0, \hat{\beta}_1$ nevyhovujú odhadu lineárneho modelu na danej hladine $1-\alpha$ [5].

Analýzu rozptylu vzhľadom k adekvátnosti modelu môžeme realizovať podľa postupu uvedeného v tabuľke č.2.

Tab.2 Postup analýzy rozptylu vzhľadom k adekvátnosti modelu

zdroj variability	odchýlka od empirickej krivky	chyba merania
súčet štvorcov	$\sum_{i=1}^n w_i (\hat{y}_i - \bar{y}_i)^2$	$\sum_{i=1}^n \sum_{j=1}^{p_i} (y_{ij} - \bar{y}_i)^2$
stupne voľnosti	$n - k - 1$	$\sum_i p_i - n$
priemerný štvorec	s_r^2	s_e^2
testovacia veličina F	$F = \frac{s_r^2}{s_e^2}$	
kvantil F	$F(1-\alpha, n-k-1, \sum p_i - n)$	
nulová hypotéza	$\eta = X \cdot \beta$	

Ak $F \geq F(1-\alpha, n-k-1, \sum p_i - n)$, potom zamietneme nulovú hypotézu, že model $\eta = X \cdot \beta$ je adekvátny k nameraným hodnotám (na hladine α).

Analýza rozptylu vzhľadom k priemernej hodnote je vykonaná nasledovne.

Tab. 3 Postup analýzy rozptylu vzhľadom k priemernej hodnote

zdroj variability	odchýlka od priemeru	odchýlka od empirickej krivky	chyba merania
súčet štvorcov	$\sum_{i=1}^n w_i (\hat{y}_i - \bar{y})^2$	$\sum_{i=1}^n w_i (\bar{y}_i - \hat{y}_i)^2$	$\sum_{i=1}^n \sum_{j=1}^{p_i} (y_{ij} - \bar{y}_i)^2$
stupne voľnosti	k	$n - k - 1$	$\sum_i p_i - n$
priemerný štvorec	s_k^2	s_r^2	s_e^2
testovacia veličina F	$F = \frac{s_k^2}{s_r^2}$		
kvantil F	$F(1-\alpha, k, \sum p_i - k - 1)$		
nulová hypotéza	$\eta = \beta_0$		

Záver

Pri riešení konkrétnych úloh je otázka príčinných či iných súvislostí len výnimočne možné zúžiť na prípad závislosti dvoch premenných. U málokterého pojmu existuje toľko rôznych odtieňov, modifikujúcich jeho význam, ako u pojmu závislosť. Ide o to, že závislosť dvoch



premenných môže byť vyjadrená funkčným predpisom, ktorý môže mať buď celkom všeobecnú platnosť, alebo môže vyhovovať konkrétnym situáciám len s určitou pravdepodobnosťou [5]. Ďalej ide o to, že samotné vzťahy medzi premennými môžu byť časovo i priestorovo odlišné a môžu mať aj jednostranný a vzájomný charakter. Ale predovšetkým ide o možnosť zámieny skrytých príčinných vzťahov za skutočnosti, ich podstatu lepšie charakterizujú pojmy súvislosť, podobnosť, či dokonca klamná korelácia.

Súhrn

Empirický prístup, založený na analýze, zhodnotení a zovšeobecnení kladných i záporných skúseností praxe. V literatúre sa zvyčajne prezentuje tak, že sa udávajú názory významných predstaviteľov empirikov, ktoré sa týkajú jednotlivých stránok riadenia. Väčšina predstaviteľov empirického prístupu sa snaží vypracovať na základe štúdia praxe riadenia konkrétne odporúčania pre manažérske konanie.

Tento článok bol vytvorený realizáciou grantového projektu VEGA č. 1/0102/11 Metódy a techniky experimentálneho modelovania vnútropodnikových výrobných a nevýrobných procesov a projektu "Centrum výskumu riadenia technických, environmentálnych a humánnych rizík pre trvalý rozvoj produkcie a výrobkov v strojárstve" (ITMS: 26220120060), na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja – aktivita 3.1 Integrované projektovanie výrobných systémov na báze fyzického a virtuálneho modelovania.

Kľúčové slová

Empirický model, analýza rozptylu, rozdelenie.

Použitá literatúra

- [1] MAROŠ, B.: Empirické modely I. – Analýza inženýrskeho procesu, CERM, Brno, 2001
- [2] BARD, Y.: Nonlinear parameter estimation, Academic press, New York 1974,
- [3] EPPINGER, S.: Model-based approaches to managing concurrent engineering. Journal of Engineering Design, vol.2 č.4,
- [4] KRÄMER W., SONNBERGER, H.: The linear regression model unter tests, Physica Verlag Heidelberg,
- [5] MELOUN, M., MILITKY, J.: Statistické spracování experimentálních dát, Edice Plus Praha 1994,
- [6] SEARLE, S. R.: Linear models, J. Wiley, New York.

Kontaktná adresa

doc. Ing. Peter Trebuňa, PhD.
Technická univerzita v Košiciach, Strojnícka fakulta
Katedra priemyselného inžinierstva a manažmentu
Němcovej 32, 040 01 Košice
e-mail: peter.trebuna@tuke.sk.