

STRUCTURE OF STATISTIC METHODS IN MECHANICAL ENGINEERING
 OR "ALMOST NOTHING ABOUT EVERYTHING"

STRUKTURA STATISTICKÝCH METOD VE STROJÍRENSTVÍ
 ANEB "TĚMĚŘ NIC O VŠEM" - ČÁST II.

C. Kratochvíl, P. Janíček

The contribution continues the first part and deals with
 basic static and dynamic statistic methods.

1. ÚVOD.

Příspěvek je pokračováním stejnojmenného předcházejícího příspěvku, ve kterém byl uveden stručný přehled vstupních a redukčních statistických analýz. V tomto příspěvku je pozornost zaměřena především na heslovitě podanou obsahovou náplň statických a dynamických statistických metod aplikovatelných na soubory dat získané měřením na technických soustavách (TS). Předpokládá se, že čtenáři jsou jasné pojmy a obsahová náplň statistické analýzy jednorozměrných dat.

2. STATISTICKÉ REALIZAČNÍ METODY.

Jsou zde postupně uvedeny analýza regresní, korelační, disperzní a kovarianční.

Regresní analýza (RA) - cílem regresní analýzy je získat a objasnit vztah mezi závisle proměnnou veličinou y a nezávislými proměnnými x_i , které mají charakter spojitých (číselných, kvantitativních) veličin. Je nutno rozlišovat tyto případy:

- rozptyly měřených hodnot veličin x_i jsou srovnatelné s rozptyly měřené veličiny y (tzv. *funkcionální modely*, [1]),
 - rozptyly veličin x_i jsou zanedbatelné vzhledem k rozptylům y (tzv. *strukturní modely*, [1]) - tento případ bude analyzován.
- Regresní závislost, vyjádřená funkcí $y = \beta(x, \beta) + e_i$, v níž β jsou tzv. *regresní koeficienty*, může být lineární vzhledem ke koeficientům β ($y = \beta_1 + \beta_2 x_1 + \beta_3 x_2$; $y = \beta_1 + \beta_2 \cos x$) nebo nelineární neseparabilní ($y = \exp(\beta_1 x) + \exp(\beta_2 x)$), separabilní ($y = \beta_1 + \exp(\beta_2 x)$), vnitřně lineární ($y = \beta^3 x$) nebo linearizovatelná ($y = \beta_1 \exp(\beta_2/x)$).

Následující text se týká RA lineární a linearizovatelné (nelinearity vzhledem k β lze převést vhodnou transformací na lineární případ).

Nejužívanější metodou k získání odhadů koeficientů β je metoda nejmenších čtverců (MNC), která je aplikovatelná za těchto předpokladů:

- 1) v matici proměnných x_i nejsou žádné dva sloupce kolineární, a dále náhodné chyby x_i ;
- 2) mají nulovou střední hodnotu,
- 3) konstantní a konečný rozptyl,
- 4) jsou vzájemně nekorelované,
- 5) mají normální rozdělení.

Důležitým krokem RA je určení konfidenčních intervalů regresních koeficientů v nichž leží teoretická hodnota β se zvolenou pravděpodobností $(1 - \alpha)$ pro zvolenou hladinu významnosti α .

Součástí RA je:

- *Statistická analýza* RA zahrnující testování významnosti odhadů β , významnosti absolutního členu β_1 , multikolinearity a správnosti navržené lineární regresní závislosti.
- *Regresní diagnostika* obsahující tyto kroky: posouzení kvality dat pomocí statistických analýz reziduí, analýz prvků projekční matice a grafů pro zjištění vlivných bodů (viz průzkumová analýza, [část I.]), posouzení kvality navržené regresní závislosti s využitím parciálních regresních a reziduálních grafů a znaménkových testů.
- *Ověření předpokladů* MNC: není-li splněn předpoklad 1) jedná se o tzv. multikolinearitu, 3) heteroskedasticitu, 4) autokorelaci, 5) porušení normality chyb. Pro tyto případy jsou rozpracovány speciální přístupy umožňující získat odhady koeficientů β_i [1].

U RA nelineární vzhledem k parametrům β vede použití libovolného kritéria regrese (metoda nejmenších čtverců, metoda maximální věrohodnosti) na úlohu extremizace, u níž "proměnnými" jsou regresní parametry. Extremalizační úlohu lze řešit s využitím obecných optimalizačních metod, a to:

- hledáním volného extrému - pokud nejsou na regresní parametry kladena žádná omezení,
- hledáním vázaného extrému - když regresní parametry musí splňovat určité omezující podmínky.

Algoritmy pro nelineární regresi lze členit do těchto základních skupin [1]:

- nederivační optimalizační postupy - patří sem metoda přímého hledání (Rosenbrockova, Powellova), simplexové metody, metody využívající náhodných čísel (algoritmy CRS, GSA, AST, ARST), postupy metody nejmenších čtverců,
- derivační metody pro kritérium shody nejmenších čtverců - jako Gaussovy-Newtonovy metody, metody Marquardtova typu,
- obecné derivační metody,
- algoritmy pro speciální případy.

Obdobně jako u lineární regrese i součástí nelineární regrese je statistická analýza odhadů regresních parametrů a regresní diagnostika [1].

Korelační analýza (KA) - cílem je vyšetření existence a kvantifikace závislosti mezi veličinami $f_1, \dots, f_m$ m -rozměrného náhodného vektoru f s využitím regrese a tzv. *korelačních koeficientů*. Prvky náhodného výběru náhodných veličin f_1, f_2 označme

x_1, x_2 . Je možno odvodit vztahy pro teoretické regrese $E(f_2/x_1)$; $E(f_1/x_2)$, které pro eliptická podmíněná rozdělení jsou lineární a v souř. systému f_1, f_2 představují přímky. Součin jejich směrnic je roven čtverci korelačního koeficientu ρ . Pro $\rho = 1$ obě regrese splnou a mezi f_1, f_2 existuje funkční závislost, pro $\rho = 0$ veličiny f_1, f_2 jsou na sobě nezávislé. Pro případ více náhodných veličin se k vyjádření jejich závislosti používají tyto koeficienty:

- *Vícendobný korelační koeficient* - $R_{1(2, \dots, m)}$ - definuje míru lineární závislosti mezi f_1 a nejlepší lineární kombinací složek $f_2, \dots, f_m$. Je-li roven jedné, pak f_1 je přesně lineární kombinací ostatních veličin x .
- *Párový korelační koeficient* R_{ij} - vyjadřuje míru lineární stochastické závislosti mezi veličinami f_i, f_j . Pozor! Významná párová korelace není důkazem příčinné souvislosti - někdy vznikají falešné korelace. Proto je nutné analýzu závislosti doplnit o tzv.
- *parciální korelační koeficient* - slouží k posouzení závislosti mezi složkami f_1 a f_j při vyloučení působení dalších složek i, j .
- *pořadový (neparametrický) korelační koeficient* - nahrazuje klasický párový koeficient, přičemž je málo citlivý na přítomnost vybočujících hodnot.

Disperzní analýza (DA) (analýza rozptylu) - cílem je získání informací, zda nezávisle proměnná diskrétního typu, tzv. *faktor*, ovlivňuje hodnotu nezávisle proměnné (kvantitativní veličina). Faktorem je buď veličina diskretizovaná z veličiny spojitě (např. úrovně teplot), nebo se jedná o veličinu alternativní (legující prvek: ano - ne) či nominální (typ metody, stroje, kvalifikace pracovníka,). DA se realizuje tak, že se provádí j -opakovaných měření na různých úrovních u_j faktoru U . Proměnná y se pak vyhodnocuje ze vztahu $y_{ij} = \mu + \alpha_i + e_{ij}$, v němž μ je celková střední hodnota ze všech měření na všech úrovních $i, \alpha_i = \mu_i - \mu$ je tzv. *efekt úrovně* u_i (μ_i je střední hodnota měření na úrovni u_i), e_{ij} je náhodná chyba. Podstata DA je v tom, že celkový rozptyl dat se rozkládá na složky objasněné (známé zdroje variability) a na složku neobjasněnou, o níž se předpokládá, že je náhodná. Pak se testuje shoda středních hodnot na jednotlivých úrovních u_j , tedy hypotéza $H_0: \alpha_i = 0$ (oproti hypotéze $H_1: \alpha_i \neq 0$). Jinými slovy se testuje, zda rozptyly měření na jednotlivých úrovních u_j jsou (nejsou) srovnatelné s rozptyly mezi jednotlivými úrovněmi u_j . Jestliže se potvrdí H_0 , pak jednotlivé úrovně u_j nemají vliv na hodnotu proměnné y . V opačném případě, jedná-li se o nezávisle proměnné diskretizované veličiny, může se pokračovat ve statistické analýze regresní analýzy, při alternativních či nominálních proměnných je možno testovat tzv. *kontrasty*, např. $H_0: \alpha_1 = \alpha_2; H_0: \alpha_1 - \alpha_2 = \alpha_3 - \alpha_4, \dots$.

Sleduje-li se vliv jednoho faktoru U na veličinu y je to tzv. *jednofaktorová DA*, při vícero faktorech *vícefaktorová DA* (kombinace jednotlivých faktorů vytvářejí pak tzv. *latinské*, či *řeckolatské čtverce*). Je možno sledovat i interakce jednotlivých úrovní jednotlivých faktorů. Dále se rozeznává DA s tzv. *pevnými efekty*, jsou-li jednotlivé úrovně faktorů pouze výběrem

z konečného souboru, či DA s *náhodnými efekty*, jedná-li se o výběr faktorů z nekonečného souboru. Statistické zpracování je u obou případů stejné, odlišné jsou jen předpoklady DA a způsob interpretace výsledků.

Kovariační analýza (K_{OA}) (analýza kovariance) - je statistická metoda u níž závisle proměnná y je kvantitativní veličina a nezávisle proměnné mají charakter jednak spojitý (x), jednak diskrétní (u). Cílem K_{OA} je posoudit a analyzovat závislost y na u , "očistěné" od vlivu x . V praxi totiž existují případy, u nichž analýza rozptylu, použitá pro určení závislosti veličiny y na úrovních u faktoru U , dává nesprávné výsledky, je-li y navíc závislé na spojitě nezávisle proměnné x . V úlohách s jednou veličinou y , jednou proměnnou x a jedním faktorem U se pracuje s matematickým vztahem $y = \mu + \alpha_1 + \beta_2$, kde μ je celková střední hodnota, α_1 je efekt úrovně u_i a β je regresní koeficient. Kovarianční analýzu lze chápat jako spojení regresní a disperzní analýzy. V procesu K_{OA} se provádí tyto statistické testy:

- shoda regresních koeficientů ve skupinách, tedy $H_0: \beta_1 = \dots \beta_k = \beta$; jestliže H_0 není potvrzena nelze v K_{OA} pokračovat standardním způsobem,
- nulovost regrese, tedy $H_0: \beta = 0$; realizace K_{OA} nemá význam a závislost y na u_i se provede analýzou rozptylu,
- ověření $H_0: \alpha_1 = \dots \alpha_i = 0$, což je hlavní náplní K_{OA} . Je-li H_0 potvrzena, pokládá se vliv úrovně u_i faktoru U na proměnnou y na zvolené hladině významnosti za podstatný při "očistění" proměnné y od spojitě proměnné x .

Jestliže v K_{OA} se vyskytuje více spojitých proměnných x jde o tzv. *vícenásobnou* K_{OA} , pokud je více proměnných y je to *více-rozměrná* kovarianční analýza. [2].

3. DYNAMICKÉ REALIZAČNÍ METODY - ANALÝZY NÁHODNÝCH PROCESŮ

Patří se metody zabývající se statistickým zpracováním těch projevů a procesů na technických soustavách (TS), jejichž parametry jsou podstatně časově závislé. Patří sem např. problematika vibrací, otřesů a hluku u TS, jejich odezev na náhodná zatěžování apod. V základním členění dynamických metod lze ho-vořit o:

- Analýza signálů - sloužící ke zjišťování charakteristik odezev TS na neznámé obecné buzení, které není předmětem našeho zájmu. Analýzu signálů lze charakterizovat jako zpracování výsledků měření vibrací prvků TS za určitých provozních podmínek takovým způsobem, aby byly snadno interpretovatelné a použitelné pro další využití, např. v simulačním modelování.
- Analýza soustav - je to analýza dynamických vlastností a chování TS, která nejen analyzuje signály odezev TS, ale vyšetřuje mezi nimi i příčinné závislosti, a to ve tvaru tzv. *vstupně-výstupních (V-V) charakteristik* TS. Tyto vyjadřují relace mezi vstupními (budícími) a výstupními (odezvami) signály na TS. K analýze soustav je k dispozici řada dynamických statistických metod, z nichž ty nejdůležitější jsou zde stručně vymezeny.

- Modální analýza (MA) - je to analýza základních vlastností kmitů TS v jejich různých stavech, jako přechodových, provozních, hraničních, mezních apod. Znalost vlastností kmitů TS je důležitá z hlediska modelování (experimentálního nebo výpočtového) dynamických odezev mechanické soustavy v tom smyslu, že libovolnou dynamickou odezvu TS lze zredukovat na diskrétní soubor tvarů kmitů (vidů), který obsahuje informace o:
 - tvarech (případně i o pořadí) jednotlivých kmitů, - hodnotách frekvencí a velikostech tlumení. V rámci MA se vyšetřuje důležitá V-V charakteristika, tzv. *přenosová funkce* $H(\Omega)$, představující relaci mezi vstupními $X(\Omega)$ a výstupními $F(\Omega)$ signály TS, danou vztahem $X(\Omega) = H(\Omega) \cdot F(\Omega)$. Funkce $H(\Omega)$ je určena svojí absolutní hodnotou $|H(\Omega)|$ a fází $\angle(\Omega)$. S využitím $H(\Omega)$ lze vyjádřit spektrum signálů na výstupu z TS na základě spektra signálů vstupních. V případě lineární dynamiky je $H(\Omega)$ nezávislou (invariantní) charakteristikou vzhledem k druhu vstupních a výstupních signálů (harmonických, impulsových, náhodných, ...), je tedy invariantní vzhledem k charakteru dynamických stavů TS (pokud zůstávají v oblasti lineární dynamiky). U lineárních modelů platí jak známo *princip superpozice*, což znamená, že jakákoliv spektra vstupních signálů lze rozdělit na parciální účinky, pro které lze vyhodnocovat dílčí odezvy a nakonec je sečíst. Analýza přenosových funkcí pak umožňuje nalézt spektra vlastních frekvencí i rezonancí.

Jedná-li se o nelineární dynamiku, pak $H(\Omega)$, určující charakter signálů na výstupu z TS, odpovídá zcela určitým vstupům - neplatí zde tedy uvedená invariantnost. Ve strukturách TS existuje nespočetné množství prvků způsobujících nelineární chování TS. Patří sem např. vůle, předpětí, hystereze, v elektrických soustavách releové charakteristiky apod. Dílčí metodiky řešení těchto soustav (např. metody ekvivalentní nebo stochastické linearizace), zvládnutelné běžně dostupnými výpočetními prostředky na vyhovující úrovni přesnosti, lze doporučit jen pro omezený počet nelinearit, zahrnující nashlé základní typy nelinearit vyskytujících se v technické praxi. U linearizovaných TS pak hovoříme o linearizované přenosové funkci $H_e(\Omega)$.

Spektrální analýza - je to statistická analýza u níž se k popisu vstupních a výstupních charakteristik používá tzv. *výkonových spektrálních hustot* $G(\Omega)$ (VSH). Pro odezvu TS lze pak psát $G_{XX} = H(\Omega) \cdot G_{FF}(\Omega) \cdot H^{*}(\Omega)$, kde $G_{XX}(\Omega)$ je VSH na výstupu, $G_{FF}(\Omega)$ je VSH na vstupu, T^{*} ... transponovaná komplexně sdružená. Vzhledem k tomu, že $G(\Omega)$ je kvadratickou funkcí w jsou významné složky frekvenčního spektra zesíleny na úkor složek méně významných. Výkonová spektra výstupních signálů se tak stávají prakticky nevýznamnější charakteristikou současné inženýrské praxe. Přispívají k tomu i dokonalé procedury FFT (Rychlá Fourierova transformace) pro jejich získání ze souboru naměřených (vypočtených) dat.

Kepstrální analýza - u této statistické metody se k popisu dynamického chování TS používá tzv. *kepstrum*, což je vlastně *výkonové spektrum logaritmu výkonové spektrální hustoty* analyzo-

vaného signálu. Matematicky vyjádřeno $C(\zeta) = |F\{\log G(f)\}|^2$. Kepstrum je citlivější než klasické výkonové spektrum při analýze určitých artefaktů, vyplývajících z výběru spektrálních parametrů a užitých výpočtových metod (např. detekce významných harmonických složek především v okrajových pásmech spekter, fázových vztahů, modulačních frekvenčních efektů apod.).

Koherenční analýza - na základě koherence lze odhadnout míru lineárního vztahu mezi signály na vstupu a na výstupu TS. Ne rovnost $|G_{XF}(\Omega)|^2 \leq G_{XX}(\Omega) \cdot G_{FF}(\Omega)$, kde $G_{XF}(\Omega)$ je vzájemná VSH vstupu a výstupu, vyjadřuje skutečnost, že jedno ze spekter obsahuje nekoherentní šumy, způsobené nelinearitou TS. Stupeň lineární závislosti pak lze odhadnout na základě *koherenční funkce*, $\zeta^2(\Omega) = |G_{XF}(\Omega)|^2 / G_{XX}(\Omega) \cdot G_{FF}(\Omega)$, přičemž jak známo $0 \leq \zeta^2(\Omega) \leq 1$. Mezní hodnota 1 odpovídá měření "bez šumu", mezní hodnota 0 odpovídá měření "jen šumu".

Kumulantní analýza - v přehledu matematických prostředků, užívaných k popisu náhodných dějů, se většinou spokojujeme se znalostí vhodných číselných charakteristik (korelačních funkcí, funkcí výkonových spektrálních hustot apod.), místo znalosti odpovídajících *zákonů rozdělení pravděpodobností procesů*. Potiže spojené s řešením vztahů obsahujících zákony rozdělení pro obecně negaussovske procesy, působících navíc v nelineárních soustavách, vedou k závěru, že je často účelné věnovat se nejen metodám výpočtů pravděpodobnostních momentů, ale i *kumulantů* (semiinvariantám), které mají v jistém smyslu význam navzájem ortogonálních číselných charakteristik. Protože mezi kumulanty a momenty existují jednoznačné, i když často velmi složité vztahy, je účelné vytvořit speciální metodický aparát tzv. "kumulantní analýzy" pro řešení některých typů nelineárních úloh. Podrobnosti jsou uvedeny ve specializované literatuře.

LITERATURA

- [1] Meloun, M.; Militký, J.: Statistické zpracování experimentálních dat na osobních počítačích. Nakladatelství FINISH, Pardubice 1992
- [2] Hebák, P.; Hustopecký, J.: Vícerozměrné statistické metody s aplikacemi. SNTL, Praha 1987.
- [3] Ondráček, E.; Janíček, P.: Výpočtové modely v technické praxi. SNTL, Praha 1990.

Prof. Ing. Přemysl Janíček, DrSc.; Ústav mechaniky těles, FS VUT v Brně, Technická 2, 616 69 Brno; tel.: 05 41142807, fax: 05 745718
e-mail: JANICEK@UMTN.FME.VUTBR.CZ

Prof. Ing. Ctirad Kratochvíl, DrSc.; Ústav mechaniky těles, FS VUT v Brně, Technická 2, 616 69 Brno; tel.: 05 41142853, fax: 05 745718
e-mail: KRATOCHVIL@UMTN.FME.VUTBR.CZ