EAN '95

33rd Conference of Experimental Stress Analysis
33. konference o experimentální analýze napětí
5. - 8.6.1995, Třešť, Czech Republic

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

STRUKTURA STATISTICKÝCH METOD VE STROJÍRENSTVÍ ANEB "TÉMĚŘ NIC O VŠEM" - ČÁST I.

P. Janíček, C. Kratochvíl, M. Druckmüller

In this and the following contribution of the same name, statistic analyses used for the experimental data processing in mechanical engineering are sorted and defined. This contribution is dedicated to input static analyses - classificatory and explorative analyses, and further to reductive static analyses. Contribution are to assist engineers for the selecting of statistic methods.

1. ÚVODNÍ SLOVO.

Čtenáři tohoto příspěvku, pokud je neodradí jeho název, se asi rozdělí do tří skupin. Ta první bude považovat za troufalost analyzovat to, o čem jsou napsány knihy, druhá ho přejde bez povšimnutí a té třetí možná poslouží k uplatnění toho co znají. První skupině se autoři omlouvají, ostatním děkují za povšimnutí.

Snahou autorů bylo vytvořit stručný a utříděný přehled statistických metod, v němž by zájemce získal první orientaci o tom jakou statistickou analýzu je vhodné použít pro zpracování souboru experimentálně zjištěných informací. Z hlediska chronologie použití statistických metod v procesu zpracování dat lze je členit na vstupní, redukční a realizační. První dvě skupiny metod jsou uvedeny v tomto příspěvku, statické a dynamické realizační metody jsou uvedeny v příspěvku následujícím.

2. VSTUPNÍ STATISTICKÉ ANALÝZY.

Zabývají se analýzou typů veličin v souboru statistických dat, posuzují se jejich vlastnosti, specifika a závislosti a přiřazuje se odpovídající statistická metoda - tento proces lze označit jako *klasifikační analýzu*. Dále se posuzuje zda jsou splněny předpoklady pro aplikaci vybrané statistické metody - to je náplní tzv. *průzkumové analýzy* [1].

2.1 **Klasifikační analýza**

souboru experimentálních dat obsahuje tyto dílčí analýzy:

1. Analýza typů veličin v souboru dat. Z hlediska členění statistických analýz je důležité členění veličin podle počtu hodnot, které veličina může nabývat, na veličiny:
   + *spojité* - nabývají nespočetně mnoha hodnot a mají především charakter veličin číselných (kvantitativních),
   + *diskrétní* - nabývají ve svém oboru konečného nebo spočetného počtu hodnot. Lze je členit na:
     - *alternativní* (binární) - nabývají jen dvou různých hodnot, např. typy pohlaví, aplikace metod (kaleno - nekaleno), kvalita výrobků (vyhovující, zmetek, ...)
     - *nominální* - nabývají více než dvou různýché hodnot, např. typy barev, typy strojů, způsoby tepelného zpracování, místa odběrů vzorků, ...)
     - *pořadové* - jsou typu nominálního s předepsaným pořadím, např. hodnocení užitečnosti, klasifikační stupnice, ...
     - *diskretizované* - veličiny vznikají výběrem konkrétních hodnot ze spojitých veličin.

2. Analýza typů závislostí mezi veličinami
   + Je-li vyšetřovaná veličina v časovém intervalu jejího experimentálního určování na ničem podstatně nezávislá a to v tom smyslu, že hodnoty fyzikálních veličin, na nichž může být závislá, jsou cíleně udržovány konstantní. Je to případ, který se ve statistice běžně označuje jako *analýza jednorozměrných dat*. Všechny ostatní typy analýz jsou součástí vícerozměrných statistických metod [2].
   + Obsahuje-li soubor experimentálních dat dvě nebo více různých fyzikálních veličin je úkolem statistické analýzy vyšetřit existenci závislosti mezi veličinami, případně je rozčlenit na nezávislé a závislé a určit míru závislosti. To je náplň *korelační analýzy* pro dvě nebo více proměnných. Určováním míry lineární závislosti lineárních funkcí dvou skupin náhodných veličin se zabývá tzv. *kanonická korelace* (kanonická analýza).
   + Je-li vyšetřovaná veličina podstatně závislá na čase, jedná se o *analýzu náhodných procesů* (jsou to dynamické statistické metody, viz II. část příspěvku).
   + Závisí-li vyšetřovaná veličina podstatně na poloze vzhledem k zavedenému souřadnicovému systému, jedná se o *analýzu polí*.
   + Závisí-li vyšetřovaná veličina podstatně na ostatních veličinách z různých jiných důvodů - příčinně, druhově, na typu procesu, apod. Podle toho jaké typy veličin se vyskytují v těchto závislostech, rozeznáváme tyto typy statistických analýz:

- *Regresivní analýzu* - závisle i nezávisle proměnné jsou veličiny spojité - jedná se o závislost příčinnou,
- *Analýzu rozptylu* (disperzní analýzu) závisle proměnná je spojitá, nezávisle proměnná je diskrétního typu - může se jednat o závislost *příčinnou* (nezávisle proměnná je diskretizovaná spojitá veličina), *typovou* (různé typy strojů, metod, přístrojů), *druhovou* (různé legující prvky), *polohovou* (různá místa odběru vzorků), *věkovou, pohlaví, vzdělanostní*, atd.

- *Kovarianční analýza* - závisle proměnná je spojitá, nezávisle proměnné jsou kombinacemi spojitých a diskrétních veličin - může se jednat o příčinně-typové závislosti apod. Uvádí se [2], že kovarianční analýza je syntézou regresní a disperzní analýzy.
- *Diskriminační analýza* - vyšetřuje se závislost jedné nebo více kvalitativních veličin (alternativních) na několika kvantitativních, obvykle spojitých, veličinách. Touto analýzou se např. určuje původ výrobku, způsob jeho zatížení na základě experimentálně zjištěných důsledků (porušení soudržnosti,chování výrobku).

Regresní, disperzní a kovariační analýzy lze ve smyslu vymezení přímých a nepřímých úloh, např. [3], označit za *přímé statistické analýzy* - nezávislé veličiny charakterizují působení na objekt, resp. jeho strukturu, určované, závislé proměnné veličiny pak experimentálně zjištěné vlastnosti a chování objektu.

Diskriminační analýzy lze považovat za nepřímé (inverzní) metody, protože nezávislé veličiny vyjadřují vlastnosti a chování objektu, závislé veličiny alternativního typu přiřazují uvedeným důsledkem působení na objekt nebo jeho typ či strukturu.

2.2 Průzkumová analýza

Posláním této analýzy je odhalit a posoudit vlastnosti a zvláštnosti souboru dat a ověřit zda jsou splněny předpoklady pro vybranou statistickou metodu dle kap. 2.

K základním statistickým zvláštnostem patří např. tyto charakteristiky:
1) stupeň symetrie rozdělení výběru,
2) jeho špičatost,
3) lokální koncentrace dat,
4) přítomnost vybočujících dat.

K posouzení těchto zvláštností u jednorozměrných dat se používají nejrůznější typy grafů (v závorce číslo odhalované zvláštnosti), např. kvantilový graf (1,3,4), diagram rozptýlení (3,4), vrubový krabicový graf (variabilita mediánu), graf polosum (1), symetrie (1), špičatosti (2), diferenční kvantilový graf (1,2), dále jsou to jádrové odhady hustoty pravděpodobnosti, histogramy, rankitové grafy (1), pravděpodobnostní grafy, grafy rozptýlení s kvantily [1].

Ověřování předpokladů pro zpracování jednorozměrných dat se skládá např. z těchto kroků:
- určování minimální velikosti výběru,
- ověřování nezávislosti prvků výběru - autokorelace,
- ověřování normality výběru tzv. směrovými testy nebo omnibus-testy,
- ověřování homogenity výběru.

3. REDUKČNÍ STATISTICKÉ METODY

Cílem těchto metod je zmenšení (redukce) počtu nezávisle proměnných v systému veličin a jeho zprůhlednění. Tato skupina metod je tvořena těmito analýzami:

- *Komponentní analýza* (analýza hlavních komponent) - metoda vychází z analýzy kovarianční matice a pokouší se nalést skryté (neměřitelné) veličiny, tzv. *hlavní komponenty*, které vysvětlují proměnnost a závislost závisle proměnných veličin. Snaží se vyjádřit původní proměnné pomocí menšího počtu nezávislých *hlavních komponent* a přiřadit jim určitou věcnou interpretaci. Proměnné nejsou v procesu řešení rozčleněny na nezávislé a závislé, jejich závislost není vysvětlována příčinnými vztahy ale působením hlavních komponent. Tyto mají maximálně reprezentovat původní proměnné tím, že mají vysvětlovat maximum jejich rozptylu, případně maximálně reprodukovat celkovou kovarianční matici výchozích proměnných.

- *Faktorová analýza* - podstatou metody je rozbor struktury vzájemných závislostí proměnných na základě předpokladu, že tyto závislosti jsou důsledkem působení určitého menšího počtu v pozadí stojících nezměřitelných veličin, tzv. *společných faktorů* [2]. Tento rozbor se využívá ke stanovení povahy a struktury společných faktorů tak, aby:
  - co nejjednodušeji objasňovaly analyzované závislosti,
  - jejich počet byl co nejmenší, přičemž se požaduje, aby soubor společných faktorů maximálně reprodukoval nediagonální prvky kovarianční matice původních proměnných.

  Faktorová analýza je považována za rozšíření komponentní analýzy, na rozdíl od ní vychází ze snahy vysvětlit závislost proměnných.

- *Kanonická analýza* se zabývá určováním intenzity lineární závislosti lineárních funkcí dvou skupin náhodných veličin. Stejně jako v komponentní nebo faktorové analýze se i zde provádí transformace souboru vzájemně korelovaných veličin na jiný soubor s menším počtem *skrytých veličin*. Ty jsou lineárními funkcemi původních proměnných a jsou založeny na analýze kovariančních struktur původních proměnných.

- *Shluková analýza* - cílem je roztřídit množinu objektů na nichž se experimentálně získal soubor informací o jejich vlastnostech a chování do několika shluků tak, aby objekty uvnitř shluku byly co nejvíce a mezi jednotlivými shluky co nejméně podobné. Tím se dosáhne zmenšení dimenze úlohy, protože místo všech původních objektů se v dalších analýzách pracuje jen s omezeným počtem reprezentativních objektů. Po výběru a určení hodnot parametrů veličin, které charakterizují chování jednotlivých objektů, je nutné rozhodnout o způsobu hodnocení jejich vzdálenosti a podobnosti při redukci jejich počtu. Např. pro hodnocení vzdáleností kvantitativních proměnných se používají Hemingovy, euklidovské nebo Čebyševovy vzdálenosti. Pro alternativní veličiny se aplikují míry podobnosti, resp. M, RR, S nebo Jaccardův koeficient. Pokud

se v datech vyskytují nominální proměnné, pak se tyto převádějí na větší počet alternativních proměnných. Nejpoužívanějším postupem ve shlukové analýze je tzv. *aglomerativní hierarchický postup*, u něhož proces rozkladu začíná od n shluků, z nichž každý obsahuje jeden objekt. Graficky lze proces shlukování vyjádřit v podobě *dendogramu* (stromu). Strom začíná s n větvemi a postupně se v každém kroku spojují dvě větve.

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

RNDr. Miloslav Druckmüller, CSc.; Ústav matematiky, FS VUT v Brně, Technická 2, 616 69 Brno. tel.:05 41142727, e-mail:
DRUCK@MAT.FME.VUTBR.CZ