4

STATISTICKÁ ANALÝZA VÍCEROZMĚRNÝCH DAT

V technické, biologické ale také lékařské praxi se často vedle informací, obsažených v náhodném skaláru ξ, vyskytují i informace obsažené v náhodném vektoru $\boldsymbol{\xi}$ s m složkami $\xi_1, ..., \xi_m$. Příklady vícerozměrných dat jsou

a) vyjádření vlastností produktů, jako jsou potraviny, oleje, slitiny atd., pomocí řady různých analytických metod,

b) hodnocení spekter pomocí poloh a ploch absorpčních pásů, sloužící k charakterizaci a identifikaci chemických sloučenin,

c) sledování složení surovin, produktů, odpadů, v závislosti na čase nebo na místě výskytu,

d) regulace jakosti na základě různých procesních proměnných,

e) stanovení charakteristiky produktu na základě měření souvisejících proměnných, např. spekter (vícerozměrná kalibrace).

Často je účelem vícerozměrné analýzy zkoumání vztahů mezi složkami náhodného vektoru. Pro tento účel se volí koncepce latentních proměnných (hlavních komponent, faktorů, kanonických proměnných) y, které jsou lineární kombinací původních proměnných x s vhodně volenými vazbami. Latentní proměnná y je kombinací m-tice sledovaných (měřených či jinak získaných) proměnných $x_1, x_2, ..., x_m$ ve tvaru

$$y = w_1 x_1 + w_2 x_2 + \ldots + w_m x_m \quad .$$

Jednotlivé vícerozměrné metody využívají různých způsobů stanovení vahových koeficientů $w_1, w_2, ..., w_m$.

V této kapitole je věnována pozornost především postupům, patřícím do oblasti průzkumové (exploratorní) analýzy a charakterizace vícerozměrných dat. Detailní popis najde čtenář v doporučované učebnici[2] k této sbírce.

4.1 Popis vícerozměrných dat

Zdrojová matice, tj. matice výchozích dat (popisující např. řadu aut různých značek), obsahuje **proměnné** v *m* sloupcích (např. obsah motoru, výkon, spotřebu paliva, hmotnost vozu, zrychlení, výšku, šířku, délku atd.) a **objekty** v *n* řádcích (např. auta různých výrobců a značek), na nichž jsou tyto proměnné (vlastnosti) měřeny. Protože měřené proměnné mají různé jednotky, a často se řádově liší, bývá zdrojová matice před zpracováním ještě upravována, *škálována*, a to buď (a) *centrováním*, kdy se od prvků sloupce odečte jejich sloupcový aritmetický průměr, nebo (b) *standardizací* čili *normováním*, kdy se prvky centrovaných sloupců ještě vydělí svou sloupcovou směrodatnou odchylkou.

Standardní statistická analýza je založena na předpokladu, že hodnoty x_{ij} tvoří *náhodný výběr.* Tento výběr je tvořen *n*-ticí řádkových vektorů $\boldsymbol{x}_i^T = (x_{i,1}, ..., x_{i,m})$, které lze chápat jako řádky zdrojové matice anebo souřadnice *n* bodů v *m*-rozměr-ném prostoru původních experimentálních proměnných. Výběr lze vyjádřit maticí rozměru ($n \times m$)

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_i^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ \vdots & & \vdots & & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m} \\ \vdots & & \vdots & & \vdots \\ x_{n,1} & \cdots & x_{n,j} & \cdots & x_{n,m} \end{bmatrix} .$$

Řádek zdrojové matice čili *i*-tý vektor $\boldsymbol{x}_i^{\mathrm{T}} = (x_{i,1}, ..., x_{i,m})$ nazýváme *objektem* (např. auto určitého typu) a můžeme ho chápat jako jeden bod v *m*-rozměrném prostoru. Tento objekt je charakterizován svými *proměnnými*, a to buď *kvantitativními,* metrickými, tj. číselnými hodnotami, nebo proměnnými *kvalitativními*, nemetrickými.

Metrické proměnné se vyskytují ve čtyřech škálách:

(a) *Proměnné v absolutní škále* mají přirozený počátek a jeden parametr měřítka, např. obsah uhlíku v %, rychlostní konstanta.

(b) *Proměnné v poměrové škále* mají zachován podíl hodnot charakteristik $c = x_2/x_1$, např. vztah vůči standardní sloučenině, vztah vůči jevu s definovaným nulovým počátkem, parametr σ v Hammettově rovnici.

(c) *Proměnné v intervalové škále* mají zachován podíl rozdílů $c = x_2 - x_1$. Jedná se o poměrovou škálu s přirozeným počátkem pro obě srovnávané hodnoty, např. poměr absorbancí indikátoru, vztažený na absorbanci nulové linie.

(d) *Proměnné v rozdílové škále* jsou vztahovány k různému počátku, např. hodnoty časových škál, stáří atd.

Nemetrické proměnné se vyskytují ve škálách:

(a) *Proměnné v nominální škále* jsou nejméně informativní. Je kvantifikována pouze rovnost nebo různost tříd. Obsahují kód, např. barvu výrobku, vyjádřenou kódem 1 až 16, rodinný stav (svobodný 1, ženatý 2, rozvedený 3, vdovec 4).

(b) *Proměnné v ordinální škále* jsou seřazené do tříd. Je definována relace větší anebo menší mezi třídami, a také kvantifikován rozdíl, např. žebříček umístění, pořadové číslo.

(c) *Proměnné v alternativní (binární) škále* vyjadřují rovnost či nerovnost vůči nějakému kritériu. Mají binární charakter, který můžeme popsat dvojicí 1 (ano), 0 (ne).

Třídu nebo **shluk objektů** chápeme jako množinu objektů se společnými nebo alespoň podobnými proměnnými, znaky (např. auta typu BMW). Blízkost či podobnost objektů posuzujeme na základě *míry vzdálenosti objektů* v *m*-rozměrném prostoru proměnných.

Mírou vzdálenosti objektů pro *kvantitativní* proměnné jsou běžně základní metriky:

Eukleidova metrika, čili *geometrická vzdálenost*, je standardním typem vzdálenosti, který je definován vztahem

$$d_{\mathrm{E}}(\boldsymbol{x}_k, \boldsymbol{x}_l) = \sqrt{\sum_{j=1}^{m} (x_{kj} - x_{lj})^2} ,$$

Hammingova metrika, čili *Manhattanská vzdálenost*, je definována vztahem

$$d_{\mathrm{H}}(\boldsymbol{x}_k, \boldsymbol{x}_l) = \sum_{j=1}^{m} |x_{kj} - x_{lj}| ,$$

zobecněná Minkowskiho metrika vztahem

$$d_{\mathrm{M}}(\boldsymbol{x}_k, \boldsymbol{x}_l) = \sqrt[n]{\sum_{j=1}^{m} |x_{kj} - x_{lj}|^n},$$

kde pro $n = 1$ jde o Hammingovu metriku a pro $n = 2$ o Eukleidovu. Čím je n větší, tím více je zdůrazňován rozdíl mezi vzdálenými objekty. Všechny tyto metriky neuvažují závislost mezi proměnnými. Zahrneme-li do vztahu pro vzdálenost i vazby mezi proměnnými, vyjádřené kovarianční maticí $\boldsymbol{C}$, dostaneme statistickou míru, zvanou *Mahalanobisova metrika*

$$d_{\mathrm{MA}}(\boldsymbol{x}_k, \boldsymbol{x}_l) = \sqrt{(\boldsymbol{x}_k - \boldsymbol{x}_l)^{\mathrm{T}} \, \boldsymbol{C}^{-1} \, (\boldsymbol{x}_k - \boldsymbol{x}_l)}.$$

Ta se společně s Eukleidovou metrikou nejvíce používá v praxi. Ve všech uvedených případech jsou si dva objekty tím bližší, čím je jejich vzdálenost menší.

Mírou podobnosti dvou objektů či proměnných x_i a x_j může být *Pearsonův párový korelační koeficient r*. Objekty jsou si tím podobnější, čím je párový korelační koeficient větší a bližší jedničce. V případě ordinální škály je analogickou mírou podobností *Spearmanův korelační koeficient.* Podobnost binárních nebo nominálních proměnných vyjadřují různé koeficienty asociace. Označíme-li počet případů negativní shody typu 0-0 písmenem *a*, počet případů s neshodou typu 1-0 písmenem *b*, počet případů s neshodou typu 0-1 písmenem *c* a počet případů s pozitivní shodou typu 1-1 písmenem *d*, můžeme definovat tyto koeficienty podobnosti:

(a) *Sokalův-Michenerův koeficient asociace*

$$S_{SM} = \frac{a + d}{a + b + c + d},$$

(b) *Russelův-Raoův koeficient asociace*

$$S_{RR} = \frac{d}{a + b + c + d},$$

(c) *Hamannův koeficient asociace*

$$S_{H} = \frac{a + d - b - c}{a + b + c + d},$$

a také lze konstruovat *obdobu korelačního koeficientu*

$$r_{B} = \frac{a\,d - b\,c}{\sqrt{(a + b)\,(c + d)\,(a + c)\,(b + d)}}.$$

Míra podobnosti mezi objekty, charakterizovanými různými typy proměnných, se vypočte jako vážený průměr jednotlivých měr podobnosti. Na základě měr podobnosti objektů se konstruují míry podobnosti mezi objekty a třídami a míry podobnosti mezi třídami. Jako nejčastější míra podobnosti se používá vzdálenost tříd $d(x_k, x_l)$. Analogicky zde užijeme způsobů vyjádření vzdálenosti objektů, protože objekt můžeme chápat jako třídu o jednom objektu. Čím větší je vzdálenost, tím menší je podobnost:

(a) *Vzdálenost nejbližšího souseda:* nejbližší jsou ty třídy či shluky, které mají nejmenší vzdálenost mezi dvěma nejbližšími objekty dvou pozorovaných tříd.

(b) *Vzdálenost nejvzdálenějšího souseda:* nejbližší jsou ty třídy či shluky, které mají nejmenší vzdálenost mezi dvěma nejvzdálenějšími objekty.

(c) *Vzdálenost mezi těžišti tříd:* nejbližší jsou ty třídy či shluky, které mají nejmenší vzdálenost mezi svými těžišti.

(d) *Vzdálenost průměrné vazby:* nejbližší jsou ty třídy či shluky, které mají nejmenší průměrnou vzdálenost mezi všemi objekty jedné a všemi objekty druhé třídy.

4.2 Obecný postup analýzy vícerozměrných dat

Postup analýzy vícerozměrných dat záleží na typu dat a na druhu požadované informace, jež se z dat má získat.
Typ dat:

Otázky: Před vlastní analýzou je třeba zodpovědět tři základní otázky:

(1) Je možné rozdělit vyšetřované proměnné na *závislé* a *nezávislé*?

(2) Kolik proměnných se uvažuje jako *závisle proměnných*?

(3) V jaké škále jsou jednotlivé proměnné měřeny, tj. *kardinální* čili číselné, *ordinální* čili pořadové nebo *nominální* čili znakové. Kardinální škála se označuje jako *metrická* a ostatní dvě, ordinální a nominální, jako škály *nemetrické*.

Odpovědi:

(1) Pokud je odpověď na první otázku kladná, volí se techniky pro stanovení *vztahu* mezi závisle proměnnými a vhodnou kombinací nezávisle proměnných.

(2) Pokud je odpověď na první otázku záporná, volí se techniky pro stanovení *vzájemných vazeb*, tj. provádí se simultánní analýza všech proměnných.

Typ informace:

Jednotlivé techniky pro *stanovení závislosti* se dále dělí podle počtu závisle proměnných a podle škály měření. Schematicky lze vztahy mezi jednotlivými technikami analýzy vícerozměrné závislosti zapsat ve formě těchto přiřazení:

(a) **Kanonická korelace (CC)**:

$$\underset{\text{(metrická, nemetrická)}}{y_1 \,\%\, y_2 \,\%\, \ldots \,\%\, y_m} \quad \mathsf{Z} \quad \underset{\text{(metrická, nemetrická)}}{x_1 \,\%\, x_2 \,\%\, \ldots \,\%\, x_m} \,,$$

(b) **Vícerozměrná analýza rozptylu (MANOVA)**:

$$\underset{\text{(metrická)}}{y_1 \,\%\, y_2 \,\%\, \ldots \,\%\, y_m} \quad \mathsf{Z} \quad \underset{\text{(nemetrická)}}{x_1 \,\%\, x_2 \,\%\, \ldots \,\%\, x_m} \,,$$

(c) **Analýza rozptylu (ANOVA)**:

$$\underset{\text{(metrická)}}{y_1} \quad \mathsf{Z} \quad \underset{\text{(nemetrická)}}{x_1 \,\%\, x_2 \,\%\, \ldots \,\%\, x_m} \,,$$

(d) **Diskriminační analýza (DA)**:

$$\underset{\text{(nemetrická)}}{y_1} \quad \mathsf{Z} \quad \underset{\text{(metrická)}}{x_1 \,\%\, x_2 \,\%\, \ldots \,\%\, x_m} \,,$$

(e) **Vícerozměrná regrese a kalibrace**:

$$\underset{\text{(metrická)}}{y_1} \quad \mathsf{Z} \quad \underset{\text{(metrická, nemetrická)}}{x_1 \,\%\, x_2 \,\%\, \ldots \,\%\, x_m} \,,$$

(f) **Analýza "conjoint"**:

$$\underset{\text{(metrická, nemetrická)}}{y_1} \quad \mathsf{Z} \quad \underset{\text{(nemetrická)}}{x_1 \,\%\, x_2 \,\%\, \ldots \,\%\, x_m} \,,$$

(g) **Strukturní rovnice**:

$$\begin{array}{ccc} y_1 & \mathsf{Z} & x_{11} \,\%\, x_{12} \,\%\, \ldots \,\%\, x_{1m} \\ y_2 & \mathsf{Z} & x_{21} \,\%\, x_{22} \,\%\, \ldots \,\%\, x_{2m} \\ & & \ldots\ldots\ldots \\ y_n & \mathsf{Z} & x_{n1} \,\%\, x_{n2} \,\%\, \ldots \,\%\, x_{nm} \\ \text{(metrická)} & & \text{(metrická, nemetrická)} \end{array}$$

Dělení od zcela obecné techniky (kanonická korelace) až k velmi speciálnímu případu (strukturní rovnice) umožňuje výběr konkrétní analýzy dat s ohledem na cíl analýzy a počet a typ závisle, resp. nezávisle proměnných.

4.3 Charakteristiky vícerozměrných náhodných veličin

Intenzita vztahu mezi proměnnými. K charakterizaci polohy *j*-té proměnné ξ_j, tj. *j*-tého sloupce zdrojové matice $\boldsymbol{X}$ se používá **střední hodnota** $E(\xi_j) = \mu_j$ a pro charakterizaci rozptýlení **rozptyl** $D(\xi_j) = \sigma_j^2$. Dále je třeba definovat *míru intenzity* vztahu mezi proměnnými ξ_i a $\xi_j, j = i$. Vhodnou charakteristikou je *druhý smíšený centrální moment*, nazývaný **kovariance** cov(ξ_i , ξ_j), definovaný vztahem

$$\mathrm{cov}(\xi_i, \xi_j) = E(\xi_i \xi_j) - E(\xi_i)\, E(\xi_j) .$$

Kovariance má vlastnosti:

a) Její znaménko ukazuje na trend stochastické vazby mezi *j*-tým a *i*-tým sloupcem matice.

b) Je v absolutní hodnotě shora ohraničená součinem $\sigma_i \sigma_j$, tj.

$$|\mathrm{cov}(\xi_i, \xi_j)| \le \sigma_i \sigma_j .$$

c) Je symetrickou funkcí svých argumentů.

d) Nemění se posunem počátku, ale změna měřítka se projeví úměrně jeho velikosti. Pro čísla a_1 , a_2 , b_1 , b_2 pak platí, že

$$\mathrm{cov}(a_1 \xi_i + b_1,\ a_2 \xi_j + b_2) = a_1 a_2 \,\mathrm{cov}(\xi_i, \xi_j) .$$

e) Pro nekorelované náhodné veličiny je $\mathrm{cov}(\xi_i, \xi_j) = 0$ a mohou nastat dva případy:

1. $E(\xi_i \xi_j) = 0$ a zároveň $E(\xi_i) = E(\xi_j) = 0$, což je případ *centrovaných ortogonálních* náhodných veličin, ne nutně nezávislých.
2. $E(\xi_i \xi_j) = E(\xi_i)\, E(\xi_j)$, což je případ *nezávislých* náhodných veličin.

f) Je *mírou intenzity lineární závislosti*.

Nevýhodou kovariance je fakt, že její hodnoty závisí na měřítku, ve kterém jsou vyjádřeny proměnné ξ_i a ξ_j. Její velikost je omezena součinem $\sigma_i \sigma_j$. Je proto přirozené provést standardizaci dělením tímto součinem. Vzniklá veličina ρ_{ij} =
$\rho(\xi_i, \xi_j)$ se nazývá *Pearsonův párový korelační koeficient*

$$D(\xi_i, \xi_j) = D_{ij} = \frac{\mathrm{cov}(\xi_i, \xi_j)}{\sigma_i \sigma_j} .$$

Je zřejmé, že párový korelační koeficient leží v rozmezí $-1 \le \rho_{ij} \le 1$. Pokud je $\rho_{ij} > 0$, jde o *pozitivně korelované* náhodné veličiny, a pokud je $\kappa_{ij} < 0$, jde o *negativně korelované* náhodné veličiny.
Pearsonův párový korelační koeficient má vlastnosti:

a) Rovnost $|\rho_{ij}| = 1$ ukazuje, že mezi ξ_i a ξ_j existuje přesně lineární vztah.

b) Pokud jsou náhodné veličiny ξ_i a ξ_j vzájemně nekorelované, je $\rho_{ij} = 0$.

c) V případě, že ξ_i a ξ_j pocházejí z vícerozměrného normálního rozdělení a $\rho_{ij} = 0$, znamená to, že jsou *vzájemně nezávislé*.

d) Platí, že i pro nelineárně závislé náhodné veličiny může být $\rho_{ij} = 0$.

e) Korelační koeficient ρ_{ii} náhodné veličiny ξ_i samotné se sebou je roven jedné.

f) Korelační koeficient je invariantní vůči lineární transformaci náhodných proměnných ξ_1 , ξ_2. Pro čísla a_1 , a_2 , b_1 , b_2 platí vztah

$$D(a_1 \xi_i + b_1,\ a_2 \xi_j + b_2) = \mathrm{sign}(a_1 a_2)\, D(\xi_i, \xi_j),$$

kde sign(x) je znaménková funkce, pro kterou platí

$$\mathrm{sign}(x) = \begin{cases} -1 & \text{pro } x < 0 \\ 0 & \text{pro } x = 0 \\ 1 & \text{pro } x > 0 \end{cases} .$$

Ověření normality. Jako při analýze jednorozměrných dat, hraje také u vícerozměr-ných výběrů hlavní roli předpoklad, zda data pocházejí z *normálního rozdělení*. Tento předpoklad usnadňuje zejména statistickou analýzu vektoru středních hodnot nebo kovarianční matice.

Podobně jako v jednorozměrném případě, existuje i zde řada testů, které jsou více či méně citlivé vůči různým typům narušení normality. Nenormalita může být například způsobena vybočujícími objekty $\boldsymbol{x}_i$ či pouze některými vybočujícími hodnotami x_{ij}. Mezi nejjednodušší metody ověřování normality patří testy založené na vícerozměrné šikmosti $g_{1,m}$ a vícerozměrné špičatosti $g_{2,m}$. V tomto případě se testuje simultánní platnost nulových hypotéz H_{01}: $g_{1,m} = 0$ a H_{02}: $g_{2,m} = m(m + 2)$.

Odhady parametrů polohy a rozptýlení. Z vícerozměrného výběru objektů o velikosti n, definovaného n-ticí m-rozměrných objektů $\boldsymbol{x}^{\mathrm{T}}_i = (x_{i,1}, ..., x_{i,m})^{\mathrm{T}}$, $i = 1, ..., n$, je možno stanovit *výběrový vektor středních hodnot* $\hat{\boldsymbol{\mu}}$ určený vztahem

$$\hat{\boldsymbol{\mu}} \quad ' \quad \frac{1}{n} \sum_{i}^{n} \boldsymbol{x}_i^{\mathrm{T}} \quad .$$

Podobně pro *odhad kovarianční matice* $\boldsymbol{S}^0$ platí rovnice

$$\boldsymbol{S}^0 \quad ' \quad \frac{1}{n} \sum_{i}^{n} (\boldsymbol{x}_i \,\&\, \hat{\boldsymbol{\mu}}) (\boldsymbol{x}_i \,\&\, \hat{\boldsymbol{\mu}})^{\mathrm{T}} \quad .$$

Míra polohy náhodného vektoru se charakterizuje pomocí *vektoru středních hodnot* $\boldsymbol{\mu}^{\mathrm{T}} \quad ' \quad [E(>_1), \ \ldots, \ E(>_m)]$, a míra rozptýlení pomocí *kovarianční matice* řádu $m \times m$

$$\boldsymbol{C} \quad ' \quad \begin{bmatrix} D(\xi_1) & \mathrm{cov}(\xi_2, \xi_1) & \text{þ} & \mathrm{cov}(\xi_i, \xi_1) & \text{þ} & \mathrm{cov}(\xi_m, \xi_1) \\ \mathrm{cov}(\xi_1, \xi_2) & D(\xi_2) & \text{þ} & \text{þ} & \text{þ} & \text{þ} \\ \text{þ} & \text{þ} & \text{þ} & \text{þ} & \text{þ} & \text{þ} \\ \mathrm{cov}(\xi_1, \xi_m) & \mathrm{cov}(\xi_2, \xi_m) & \text{þ} & \mathrm{cov}(\xi_i, \xi_m) & \text{þ} & D(\xi_m) \end{bmatrix} .$$

Místo kovarianční matice se používá také její normovaná verze, tj. *korelační matice*

$$\boldsymbol{R} \quad ' \quad \begin{bmatrix} 1 & \mathrm{k}_{21} & \text{þ} & \mathrm{k}_{i1} & \text{þ} & \mathrm{k}_{m1} \\ \mathrm{k}_{12} & 1 & \text{þ} & \text{þ} & \text{þ} & \text{þ} \\ \text{þ} & \text{þ} & \text{þ} & \text{þ} & \text{þ} & \text{þ} \\ \mathrm{k}_{1m} & \mathrm{k}_{2m} & \text{þ} & \mathrm{k}_{im} & \text{þ} & 1 \end{bmatrix} .$$

Korelační matice má na diagonále samé jedničky a mimodiagonální prvky jsou jednotlivé *Pearsonovy párové korelační koeficienty*. Kovarianční matice $\boldsymbol{C}$ i korelační matice $\boldsymbol{R}$ jsou symetrické.

Pro vektor výběrových středních hodnot platí

$$E(\hat{\boldsymbol{\mu}}) \quad ' \quad \boldsymbol{\mu} \qquad \wedge \qquad D(\hat{\boldsymbol{\mu}}) \quad ' \quad \frac{1}{n} \boldsymbol{C} \quad .$$

Odhad $\hat{\boldsymbol{\mu}}$ je tedy nevychýlený. Pro odhad kovarianční matice platí, že

$$E(\boldsymbol{S}^0) \quad ' \quad \frac{n \& 1}{n} \boldsymbol{C}$$

a jde o vychýlený odhad. Proto se používá *výběrová korigovaná kovarianční matice*

$$\boldsymbol{S} \quad ' \quad \frac{n}{n \,\&\, 1} \boldsymbol{S}^0 \quad ,$$

která je již nevychýleným odhadem kovarianční matice $\boldsymbol{C}$. Matice $\boldsymbol{S}^0$ je *výběrová kovarianční matice*. Odhady $\hat{\boldsymbol{\mu}}$ a $\boldsymbol{S}^0$ jsou maximálně věrohodné, tzn. že náhodný výběr, charakterizovaný maticí $\boldsymbol{X}$ pochází z normálního rozdělení $N(\boldsymbol{\mu}, \boldsymbol{C})$. Za stejných podmínek má $\hat{\boldsymbol{\mu}}$ rozdělení $N(\boldsymbol{\mu}, \boldsymbol{C}/n)$.

Pokud máme dva vektory, $\boldsymbol{\xi}_1$ a $\boldsymbol{\xi}_2$, které jsou nezávislé a stejně rozdělené se střední hodnotou $\boldsymbol{\mu}$ a kovarianční maticí $\boldsymbol{C}$, je *vícerozměrná šikmost* dána vztahem

$$g_{1,m} \quad ' \quad E[(>_1 \,\&\, \boldsymbol{\mu})^{\mathrm{T}} \, \boldsymbol{C}^{\&1} \, (>_2 \,\&\, \boldsymbol{\mu})]^3$$

a pro *vícerozměrnou špičatost* platí

$$g_{2,m} \quad ' \quad E[(>_1 \,\&\, \boldsymbol{\mu})^{\mathrm{T}} \, \boldsymbol{C}^{\&1} \, (>_1 \,\&\, \boldsymbol{\mu})]^2 \quad .$$

K vyjádření funkcí $g_{1,m}$ a $g_{2,m}$ lze využít i vícerozměrných centrálních momentů. Speciálně pro případ vícerozměrného normálního rozdělení pak platí, že

$$g_{1,m} \quad ' \quad 0 \quad \mathrm{a} \quad g_{2,m} \quad ' \quad m\,(m \,\%\, 2) \quad .$$