

ČESKÁ TECHNICKÁ NORMA

ICS 03.120.30 **Srpen 2012**

Statistická interpretace dat –
Část 4: Detekce a ošetření odlehlých hodnot

ČSN
ISO 16269- 4
01 0233

Statistical interpretation of data – Part 4: Detection and treatment of outliers

Interprétation statistique des données – Partie 4: Détection et traitement des valeurs aberrantes

Tato norma je českou verzí mezinárodní normy ISO 16269- 4:2010. Překlad byl zajištěn Úřadem pro technickou normalizaci, metrologii a státní zkušebnictví. Má stejný status jako oficiální verze.

This standard is the Czech version of the International Standard ISO 16269- 4:2010. It was translated by Czech Office for Standards, Metrology and Testing. It has the same status as the official version.

Národní předmluva

Zaměření přejímané mezinárodní normy

Jakákoliv pozorování a měření uskutečňovaná v rámci experimentů, přijímacích zkoušek, ověřovacích zkoušek, studií zaměřených na analýzu výrobního procesu atd. jsou vždy vystavena přítomnosti chyb vznikajících při odběrech dílčích vzorků, vzorků určených k získání informací o kvalitě přejímaných dávek, o způsobilosti a výkonnosti výrobního procesu, o stupni heterogenity (resp. stupni požadované homogenity) přejímaného hromadného materiálu apod. Přítomnost chyb ve všech takových datech při jejich zpracování způsobuje nejen znehodnocení často nákladných zkoušek, ale může vést i k výraznému zkreslení informací, které jsou pro dané procesy vstupními technologickými parametry. Přitom v celém souboru získaných dat může jít o přítomnost jedné nebo jen několika málo hodnot, které v praxi obvykle nazýváme odlehlými hodnotami. Předložená mezinárodní norma, která vychází v prvním vydání, nabízí nejen propracované postupy umožňující detekci přítomnosti těchto odlehlých dat, ale i postupů zaměřených na jejich ošetření. Norma je tedy určena nejen pro oblast laboratoří, ale bezpochyby najde své uplatnění na všech stupních managementu kvality.

Související ČSN

ISO 3534-1:2006 zavedena v ČSN ISO 3534-1:2010 (01 0216) Statistika – Slovník a značky – Část 1: Obecné statistické termíny a termíny používané v pravděpodobnosti

ISO 3534-2:2006 zavedena v ČSN ISO 3534-2:2010 (01 0216) Statistika – Slovník a značky – Část 2: Aplikovaná statistika

ISO 5479:1997 zavedena v ČSN ISO 5479:1998 (01 0239) Statistická interpretace údajů – Testy

odchýlení od normálního rozdělení

ISO 5725-2:1994 zavedena v ČSN ISO 5725-2:1997 (01 0251) Přesnost (správnost a shodnost) metod a výsledků měření – Část 2: Základní metoda pro stanovení opakovatelnosti a reprodukovatelnosti normalizované metody měření

Vypracování normy

Zpracovatel: Ing. Vratislav Horálek, DrSc., IČ 15949800

Technická normalizační komise: TNK 4 Aplikace statistických metod

Pracovník Úřadu pro technickou normalizaci, metrologii a státní zkušebnictví: Ing. Petr Svoboda

MEZINÁRODNÍ NORMA

Statistická interpretace dat ISO 16269- 4
Část 4: Detekce a ošetření odlehlých hodnot První vydání
2010-10-15

ICS 03.120.30

Obsah

Strana

Úvod 7

1 Předmět normy 8

2 Termíny a definice 8

3 Značky 15

4 Odlehlé hodnoty v jednorozměrných datech 16

4.1 Obecně 16

4.1.1 Co je odlehlá hodnota? 16

4.1.2 Jaké jsou příčiny odlehlých hodnot? 16

4.1.3 Proč mají být odlehlé hodnoty detekovány? 16

4.2 Třídění dat 17

4.3 Testy pro zjišťování odlehlých hodnot 19

4.3.1 Obecně 19

4.3.2 Výběr z normálního rozdělení 19

4.3.3	Výběr z exponenciálního rozdělení	21
4.3.4	Výběry odebrané z některých známých nenormálních rozdělení	23
4.3.5	Výběr odebraný z neznámých rozdělení	24
4.3.6	Cochranův test pro odlehlý rozptyl	26
4.4	Grafický test odlehlých hodnot	27
5	Metody pro jednorozměrná data s odlehlými hodnotami	28
5.1	Robustní analýza dat	28
5.2	Robustní odhady polohy	28
5.2.1	Obecně	28
5.2.2	Uřezaný průměr	28
5.2.3	Tukeyho bikvadraticky vážený M-odhad	29
5.3	Robustní odhady rozptylů	29
5.3.1	Obecně	29
5.3.2	Absolutní hodnota párové odchylky medián-medián	29
5.3.3	Bikvadraticky vážený M-odhad měřítka	30
6	Odehlé hodnoty ve vícerozměrných a regresních datech	30
6.1	Obecně	30
6.2	Odehlé hodnoty ve vícerozměrných datech	30
6.3	Odehlé hodnoty v lineární regresi	32
6.3.1	Obecně	32

6.3.2	Lineární regresní modely	32
6.3.3	Detekce odlehlých pozorování Y	34
6.3.4	Identifikace odlehlých pozorování X	34
6.3.5	Detekce vlivných pozorování	35
6.3.6	Robustní regresní postup	36
Příloha A	(informativní) Algoritmus pro postup GESD pro detekci odlehlých hodnot	38
Příloha B	(normativní) Kritické hodnoty testových statistik pro odlehlé hodnoty pro exponenciální	

Příloha C (normativní) Hodnoty součinitelů pro modifikovaný krabicový diagram 46

Příloha D (normativní) Hodnoty korekčních součinitelů pro robustní odhady parametru měřítka 49

Příloha E (normativní) Kritické hodnoty Cochranovy testové statistiky 50

Příloha F (informativní) Strukturovaný pokyn k detekci odlehlých hodnot v jednorozměrných datech 53

Bibliografie 55

Odmítnutí odpovědnosti za manipulaci s PDF souborem

Tento soubor PDF může obsahovat vložené typy písma. V souladu s licenční politikou Adobe lze tento soubor tisknout nebo prohlížet, ale nesmí být editován, pokud nejsou typy písma, které jsou vloženy, používány na základě licence a instalovány v počítači, na němž se editace provádí. Při stažení tohoto souboru přejímají jeho uživatelé odpovědnost za to, že nebude porušena licenční politika Adobe. Ústřední sekretariát ISO nepřijímá za její porušení žádnou odpovědnost.

Adobe je obchodní značka „Adobe Systems Incorporated“.

Podrobnosti o softwarových produktech použitých k vytvoření tohoto souboru PDF lze najít ve Všeobecných informacích, které se vztahují k souboru; parametry, na jejichž základě byl PDF soubor vytvořen, byly optimalizovány pro tisk. Soubor byl zpracován s maximální péčí tak, aby ho členské organizace ISO mohly používat. V málo pravděpodobném případě, že vznikne problém, který se týká souboru, informujte o tom Ústřední sekretariát ISO na níže uvedené adrese.



DOKUMENT CHRÁNĚNÝ COPYRIGHTEM

© ISO 2010

Veškerá práva vyhrazena. Pokud není specifikováno jinak, nesmí být žádná část této publikace reprodukována nebo používána v jakékoliv formě nebo jakýmkoliv způsobem, elektronickým nebo mechanickým, včetně fotokopíí a mikrofilmů, bez písemného svolení buď od organizace ISO na níže uvedené adrese, nebo od členské organizace ISO v zemi žadatele.

ISO copyright office

Case postale 56 · CH-1211 Geneva 20

Tel. + 41 22 749 01 11

Fax + 41 22 749 09 47

E-mail copyright@iso.org

Web www.iso.org

Published in Switzerland

Předmluva

ISO (Mezinárodní organizace pro normalizaci) je celosvětovou federací národních normalizačních orgánů (členů ISO). Práce na tvorbě mezinárodních norem obvykle provádějí technické komise ISO. Každý člen ISO, zajímající se o předmět, pro který byla vytvořena technická komise, má právo být v této komisi zastoupen. Práce se zúčastňují také vládní i nevládní mezinárodní organizace, s nimiž ISO navázala pracovní styk. ISO úzce spolupracuje s Mezinárodní elektrotechnickou komisí (IEC) ve všech otázkách elektrotechnické normalizace.

Mezinárodní normy se navrhují v souladu s pravidly uvedenými v části 2 Směrnic ISO/IEC.

Hlavním úkolem technických komisí je připravit mezinárodní normy. Návrhy mezinárodních norem přijaté technickými komisemi se rozesílají členským orgánům k hlasování. Zveřejnění mezinárodní normy vyžaduje schválení alespoň 75 % hlasujících členů.

Upozorňuje se na možnost, že některé z prvků tohoto dokumentu mohou být předmětem patentových práv. ISO nelze činit odpovědnou za identifikování libovolného patentového práva nebo všech patentových práv.

ISO 16269-4 byla připravena technickou komisí ISO TC 69 *Aplikace statistických metod*.

ISO 16269 sestává z následujících částí pod společným obecným názvem *Statistická interpretace dat*:

- Část 4: *Detekce a ošetření odlehlých hodnot*
- Část 6: *Stanovení statistických tolerančních intervalů*
- Část 7: *Medián - Odhad a konfidenční intervaly*
- Část 8: *Stanovení předpovědních intervalů*

Úvod

Identifikace odlehlých hodnot je jedním z nejstarších problémů při interpretování dat. Příčiny odlehlosti hodnot zahrnují chybu měření, chybu vzorkování, podhodnocení nebo nadhodnocení sdělovaných výsledků vzorkování, nesprávné zaznamenávání dat, nesprávné přiřazování pravděpodobnostních rozdělení a modelů v předpokladech týkajících se množin dat a řídce se vyskytujících hodnot pozorování atd.

Odlehlé hodnoty mohou zkreslit a znehodnotit informace obsažené v datech ze zdrojů nebo generátorů. Ve zpracovatelském průmyslu existence odlehlých hodnot narušuje efektivitu jakýchkoliv návrhů procesů/produktů a postupů řízení kvality. Možné odlehlé hodnoty nejsou nutně *chyby* nebo *omyly*. V některých situacích může odlehlá hodnota poskytnout důležitou informaci, a tedy by měla být identifikována pro další studium.

Studium a detekce odlehlých hodnot z procesů měření vede k lepšímu porozumění procesům a ke vhodné analýze dat, která následně vyústí v závěry přinášející zlepšení.

Vzhledem k tomu, že na téma odlehlé hodnoty existuje značně rozsáhlá literatura, je velmi důležité pro mezinárodní organizace určit a normalizovat rozumné skupiny metod použitelné k identifikaci a ošetření odlehlých hodnot. Začlenění této části ISO 16269 umožní obchodu a průmyslu přijmout analýzy dat prováděné členskými zeměmi a organizacemi.

Tato norma nabízí šest příloh. Příloha A poskytuje algoritmus pro výpočet testové statistiky a kritických hodnot postupu při detekci odlehlých hodnot v souborech dat získaných z normálního rozdělení. Přílohy B, D a E poskytují tabulky potřebné pro použití doporučených postupů. Příloha C poskytuje tabulky a statistickou teorii, která je základem pro konstrukci modifikovaných krabicových diagramů pro detekci odlehlých hodnot. Příloha F poskytuje strukturovaný návod a vývojový diagram k postupům doporučeným v této části ISO 16269.

1 Předmět normy

Tato norma stanovuje podrobné popisy spolehlivých postupů statistického testování a grafických metod analýzy dat pro detekci odlehlých hodnot v datech z procesů měření. Doporučuje spolehlivé a robustní postupy odhadování a postupy testování vhodné za přítomnosti odlehlých hodnot.

Tato část ISO 16269 je navržena především pro detekci a prokázání odlehlé(-lých) hodnoty(-not) z jednorozměrných dat. Pro vícerozměrná a regresní data poskytuje tato část ISO 16269 návod k postupu řešení.

Konec náhledu - text dále pokračuje v placené verzi ČSN.