Pokroky matematiky, fyziky a astronomie

Josef Machek

Kriteria hrubých chyb měření

Pokroky matematiky, fyziky a astronomie, Vol. 1 (1956), No. 3, 219--226

Persistent URL: http://dml.cz/dmlcz/137133

POKROKY
MATEMATIKY, FYSIKY A ASTRONOMIE
ROČNÍK I • ČÍSLO 3

J. MACHEK

KRITERIA HRUBÝCH CHYB MĚŘENÍ

V tomto článku bude podán přehled testů významnosti pro pozorování, odtržená od skupiny ostatních dat. V anglické literatuře jsou taková pozorování výstižně nazývána „outliers“ nebo „outlying observations“, česky jim říkáme „odlehlá pozorování“. Po prvé se tyto otázky vyskytly při astronomických, geodetických a fysikálních měřeních, pak při stanovení přesnosti chemických analys a potom upoutaly pozornost statistiků tyto testy samy o sobě, nezávisle na oboru, ve kterém jich bude užito. Přehled bude uspořádán podle historického vývoje. Historie začíná u zkoumání přesnosti měření, u ověřování, zda se v řadě měření nevyskytly hrubé chyby. K tomuto počátku celé theorie se vztahuje název článku. Budeme hovořit stále jen o aplikaci na chyby měření, i když některé z uvedených testů se hodí i k jiným účelům. Výčet existujících method bude jistě neúplný, protože některé původní prameny — zvláště starší — jsou těžko dostupné a máme k disposici jen nepřímé informace „z druhé ruky“, ze zmínek v pozdějších pracích a pod.

Začneme podrobným popisem situace, která přivedla k testům odlehlých pozorování. Je odedávna známo, že opakovaná měření téže veličiny, při nichž se usiluje o co největší přesnost, dají vždy výsledky, jež se vzájemně liší. Jsou tedy jednotlivá měření zatížena chybami. Běžně je přijato třídění těchto chyb na dvě skupiny: První skupinu tvoří chyby systematické (sistěmatičeskije pogrešnosti, constant errors, regular errors). Tyto chyby mají stejnou hodnotu při všech měřeních téže veličiny, nebo závisí funkčně na podmínkách, za kterých se měření provádí (na př. chyby způsobené refrakcí, aberací, vlivem teploty nebo vlhkosti vzduchu, při použití jednoho přístroje nesprávným dělením stupnic na přístroji, jednostranností pozorovatele). Jsou-li tyto chyby vyloučeny výpočtem oprav nebo zlepšením měřicí techniky, zbývají chyby druhé skupiny, chyby náhodné (slučajnyje pogrešnosti, irregular errors, accidental errors). Jsou to chyby způsobené na př. velmi jemným kolísáním povětrnostních vlivů, hlavně však nepravidelností v manipulaci s přístroji, vyvolanou lidským činitelem — nedokonalostí, nepřesností smyslů. Měření, na která působí jen náhodné chyby, mají ve většině případů povahu náhodných veličin, jejichž rozdělení je normální, se středem ve skutečné hodnotě měřené veličiny.

Označíme-li výsledek měření x, skutečnou hodnotu měřené veličiny μ a směrodatnou odchylku měření σ, pak tedy

$$P\{x \leqq a\} = \int_{-\infty}^{a} (2\pi)^{-\frac{1}{2}} \sigma^{-1} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] dt = \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Předpokládáme, že opakovaná měření jsou na sobě nezávislá, čili

$$P\{x_i \leqq a_i, i = 1, 2, \ldots, n\} = \prod_{i=1}^{n} P\{x_i \leqq a_i\} = \prod_{i=1}^{n} \Phi\left(\frac{a_i - \mu}{\sigma}\right).$$

Jinými slovy: řada n měření veličiny μ je náhodným výběrem rozsahu n z normálního rozdělení se střední hodnotou μ a se směrodatnou odchylkou σ. Chyby, odpovídající těmto měřením, $Z_i = x_i - \mu$, mají normální rozdělení s průměrem 0 a se směrodatnou odchylkou σ.

Nejlepšími odhady veličin μ a σ jsou maximálně hodnověrné odhady $\bar{x}$ a s,

$$x = \frac{1}{n} \sum_{i=1}^{n} x_i, \; s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2,$$

$\bar{x}$ je nestranným odhadem veličiny μ, s není nestranným odhadem σ.

Je-li μ a σ určeno dosti přesně, lze stanovit očekávanou hodnotu počtu v kterémkoli intervalu. Bylo však pozorováno, že větších chyb se vyskytuje více, než se očekává podle normálního rozdělení. To neukazuje vždy na to, že by neplatil normální zákon chyb, nýbrž spíše na to, že existují další chyby, nezachycené v klasifikaci chyb na náhodné a systematické. Tyto chyby, způsobující, že jedno nebo několik pozorování v řadě se odchýlí od skutečné hodnoty mnohem více než ostatní, se nazývají **hrubé chyby** (promachi, grubyje pogrešnosti, gross errors, blunders). Jsou způsobeny náhlým, mimořádným výkyvem v podmínkách měření (hrubým otřesem přístroje, nesprávným provedením některého úkonu, špatným přečtením hodnoty na stupnici, atd.), nějakým neobvyklým rušivým vlivem. Zahrnutí pozorování, zatíženého hrubou chybou, do odhadu $\bar{x}$ nebo s ovšem vzdálí odhady od skutečných hodnot. Proto se odjakživa pozorování, nápadně vzdálená od ostatních, vyřazovala, při výpočtu odhadů se k nim nepřihlíželo. Dálo se tak nejprve „od oka“, bez objektivního kriteria, podle subjektivního úsudku toho, kdo měření zpracovával. Ten posoudil, zda dané měření je za normálních podmínek možné, či ne. Je jisté, že někdy lze rozhodnout, zda daná hodnota je taková, že vůbec nemůže normálně vzniknout, někdy zase lze dodatečně zjistit, zda při provádění daného měření bylo všechno v pořádku. Někdy takové zjištění není možné a pro takové případy byla navrhována různá kriteria, umožňující na základě naměřených hodnot samých rozhodnout, zda extrémní hodnota má či nemá být ve výběru ponechána. Tato kriteria byla vesměs založena na počtu pravděpodobnosti.

Snad nejstarší takové kriterium navrhl prof. B. Peirce r. 1852 [1]. Jeho uveřejnění vyvolalo mezi astronomy výměnu názorů, která trvala celá desetiletí — ještě r. 1920 vyšlo odsouzení Peirceova způsobu (Stewart v *Popular Astronomy* [2]). Dnes se zdá Peirceovo kriterium zapomenuto. Jeho aplikace je pracná a v odvození se vyskytují některé dosti nejasné kroky.

Návrh jiného způsobu uveřejnil r. 1863 Chauvenet [3]. Jeho methodu lze stručně popsat takto: pravděpodobnost, že absolutní hodnota chyby $|Z|$ překročí hodnotu $t\,\sigma$, je

$$P\{|Z| > t\,\sigma\} = 2\{1 - \Phi(t)\}.$$

Matematická naděje počtu chyb, v absolutní hodnotě větších než $t\,\sigma$ v řadě n měření tedy je

$$2n\{1 - \Phi(t)\}.$$

Snadno lze najít hodnotu t_0 takovou, že tato matematická naděje je rovna 1/2. Hodnota t_0 je řešením rovnice

$$\Phi(t_0) = \frac{4n - 1}{4n}.$$

Chauvenet navrhuje zamítat chyby větší než $t_0 \sigma$ s odůvodněním, že za normálních podmínek je pravděpodobnost výskytu takové chyby menší než pravděpodobnost, že k takové chybě nedojde. Na rozdíl od Peirceova kriteria zůstalo kriterium Chauvenetovo v používání dosti dlouho, na př. v knize prof. Malikova [4] „Osnovy metrologii“ z r. 1947 je uvedeno mezi pomůckami k odhalování hrubých chyb. Proti nejjednodušším kriteriím, která stanoví na př., že se má zamítnout pozorování, vzdálené od průměru o více než trojnásobek směrodatné odchylky, má Chauvenetovo kriterium tu přednost, že stanoví hranici „přípustných chyb“ v závislosti na počtu měření. Jeho hlavní závadou však je, že závisí na parametrech μ a σ, které jsou neznámé (jinak by nebylo třeba podnikat vůbec měření). Přikazuje totiž zamítat pozorování, zatížená větší chybou než $t_0 \sigma$. Chyby se při aplikaci nahrazují odchylkami $x_i - \bar{x}$ a σ se nahrazuje svým odhadem s. Při odvození kriteria se však s takovými záměnami nepočítalo. Mimo to nevíme, jaká je pravděpodobnost „nespravedlivého zamítnutí dobrého pozorování“. Můžeme říci jen tolik: protože střední hodnota počtu chyb, v absolutní hodnotě větších než $t_0 \sigma$, je 1/2, připadá průměrně na dva výběry jedna taková chyba, i když nepůsobí žádný neobvyklý vliv. To jest, průměrně z každých dvou řad měření bude nespravedlivě vyřazeno jedno měření. (Ani tento závěr však neplatí zcela přesně, právě proto, že hodnoty parametrů byly nahrazeny svými odhady.)

Pozdější kriteria — navržená v tomto století — využívají již některých pojmů z theorie testování (ověřování) statistických hypothes. Úloha vylučování hrubých chyb je v těchto pojmech vyjádřena takto: má se ověřit hypothesa, že n nezávislých náhodných veličin má normální rozdělení se stejnými parametry, proti alternativám, že veličina, u které byla pozorována největší (nebo nejmenší) hodnota, má rozdělení s jinými parametry. Alternativy jsou takto voleny proto, že rozpaky budící pozorování budou vždy na okrajích řady. Právě vzhledem k této formulaci alternativ, která nutně vede k potížím s uspořádanými výběry, se nepodařilo založit test na standardních methodách, jako je třeba methoda poměru hodnověrnosti. Vždy se postupovalo tak, že byla navržena nějaká charakteristika (funkce pozorovaných hodnot) D jako míra vzdálenosti „podezřelého“ pozorování od ostatních. Pak byla stanovena distribuční funkce této charakteristiky za předpokladu, že všechna pozorování jsou náhodnými veličinami, jež mají normální rozdělení se stejnými průměry a rozptyly, tedy za předpokladu, že při žádném měření nezasáhne rušivý vliv. Je-li tato distribuční funkce známa, lze určit ke každému α, $0 < \alpha < 1$, takovou hodnotu $D(\alpha)$, že

$$P\{D > D(\alpha)\} = \alpha.$$

Stanoví se pak konvencí nějaká dosti malá hodnota α (na př. 0,05 nebo 0,01) a přijme se pravidlo: „podezřelé“ pozorování se z výběru vyřadí, jakmile zvolená míra jeho odchylky od ostatních dat nabude hodnoty větší nebo rovné $D(\alpha)$. V tom spočívá test hypothesy, že příslušné pozorování není zatíženo hrubou chybou. Berou-li se pak výběry rozsahu n, t. j. provádějí-li se serie n měření a nepůsobí rušivý vliv, je pravděpodobnost nespravedlivého zamítnutí jednoho pozorování rovna α. Číslo α se nazývá **hladinou významnosti**. Princip je tedy podobný, jako u starších kriterií: vyřazovat pozorování, jež jsou za normálních podmínek málo pravděpodobná, jejichž kombinace s ostatními pozorováními jsou málo pravděpodobné. Zlepšení spočívá v tom, že neurčitá fráze „málo pravděpodobná“ byla nahrazena přesným stanovením; víme nyní, jak málo pravděpodobná seskupení vyřazujeme, jak často vyřadíme pozorování, jež vzniklo za normálních podmínek a jež tedy mělo být podrženo. To umožňují přesně stanovená rozdělení kriterií D.

Při záznamu jednotlivých kriterií, která byla za míru odchylky, D, navržena, budeme užívat k označení pozorování, seřazených podle velikosti, indexů v závorkách. Zatím co x_1 značilo hodnotu, naměřenou při prvním měření, bude $x_{(1)}$ značit nejmenší pozorování, podobně $x_{(n)}$ největší, atd.

Roku 1925 podnítily některé úvahy Pearsonovy a starší práce Galtonova [5] Irwina [6] k tomu, aby užil k testu odchylky $x_{(n)}$ od ostatních pozorování rozdílu $x_{(n)} - x_{(n-1)}$. Irwin odvodil rozdělení veličin

$$\frac{x_{(n)} - x_{(n-1)}}{\sigma}, \quad \frac{x_{(n-1)} - x_{(n-2)}}{\sigma}, \quad \frac{x_{(2)} - x_{(1)}}{\sigma}, \quad \frac{x_{(3)} - x_{(2)}}{\sigma}$$

a tabeloval některé hodnoty jejich distribučních funkcí.

Podobně jako inspirovaly Galtonova a Pearsonova práce Irwina, inspirovaly Tipettovy a Pearsonovy výsledky ([7], [8]) W. Gosseta, aby použil výběrového rozpětí

$$w_n = [x_{(n)} - x_{(1)}]/\sigma$$

k ověření přesnosti chemických analys. Gosset navrhl tento postup k získání analys bez hrubých chyb [9]: provede se stanovený počet n analys. Vypočte se charakteristika w_n. Jestliže $w_n > w_n(\alpha)$,[1]) vyřadí se výsledek, který je nejvíce vzdálen od ostatních a provedou se dvě nové analysy. Znovu se počítá w_{n+1}. Je-li $w_{n+1} > w_{n+1}(\alpha)$, vyřadí se nejvzdálenější výsledek a testuje se zbylých n pozorování. Tak se postupuje, dokud nezískáme výběr alespoň n hodnot, jehož rozpětí je menší než $w(\alpha)$. Takovým vylučovacím a nahrazovacím postupem ovšem číslo α pozbude svého původního smyslu jako hladina významnosti.

Roku 1935 navrhl Mc Kay další kriterium hrubých chyb [10]. Při něm se jako míry odchylky užívá vzdálenosti pozorování od aritmetického průměru všech hodnot,

$$\mu_n = \frac{x_{(n)} - \bar{x}}{\sigma}.$$

Mc Kay odvodil rekurentní vztah mezi rozdělením této charakteristiky ve výběrech rozsahu n a $n - 1$ a našel přibližné výrazy pro hodnoty $\mu_n(\alpha)$, pro které

$$P\{\mu_n > \mu_n(\alpha)\} = \alpha.$$

Všechna tři uvedená kriteria mají stejný nedostatek: užívá se při nich směrodatné odchylky σ, která je zpravidla neznáma. Kriterium, zbavené tohoto nedostatku, navrhl r. 1935 W. R. Thompson. V práci [11] ukázal toto: zvolíme-li pevný index i, pak veličina

$$\tau_i = (x_i - \bar{x})/s$$

má rozdělení, nezávislé na parametrech μ a σ, které lze snadno převést na Studentovo rozdělení s $n - 2$ stupni volnosti. Dá se tedy bez nesnází určit hodnota τ', pro kterou

$$P\{|\tau_i| > \tau'\} = P,$$

kde P je libovolné číslo. Střední hodnota počtu hodnot τ_i v absolutní hodnotě větších než τ' ve výběru rozsahu n je nP. Přijmeme-li pravidlo zamítat pozorování x_i, pro která

[1]) $w_n(\alpha)$ je taková hodnota, pro kterou $P\{w_n - w_n(\alpha)\} = \alpha$.

$|\tau_i| > \tau'$, pak střední hodnota počtu pozorování, nespravedlivě zamítnutých z výběru rozsahu n, je nP. Jedno bezdůvodně, nespravedlivě vyřazené pozorování tedy připadá na $1/(nP)$ výběrů. Číslo $(nP)^{-1}$ zde nahrazuje hladinu významnosti.

Roku 1936 tabelovali E. S. Pearson a Chandra-Sekhar [12] přesné 100 α procentní body kriteria

$$\tau_{(n)} = \frac{x_{(n)} - \bar{x}}{s}, \left(\text{resp. } \frac{\bar{x} - x_{(1)}}{s}\right),$$

a to zajímavým postupem, aniž by odvozovali přesné rozdělení maximálních nebo minimálních hodnot $\tau_{(1)}$ nebo $\tau_{(n)}$. Existují totiž algebraické hranice hodnot $\tau_{(i)}$. Zejména

$$\tau_{(2)} \geqq -\sqrt{(n-2)/2} \quad \text{ve všech výběrech}$$

$$\tau_{(n-1)} \leqq \sqrt{(n-2)/2} \quad \text{ve všech výběrech.}$$

Tedy ze všech τ_i hodnoty menší než $-\sqrt{\dfrac{n-2}{2}}$ může nabýt jedině $\tau_{(1)}$, a hodnoty větší než $\sqrt{\dfrac{(n-2)}{2}}$ jedině $\tau_{(n)}$. Tedy rozdělení τ_i pro obecný index i závisí ve svých okrajích jen na $\tau_{(1)}$ a $\tau_{(n)}$.

Je-li tedy $\tau' > \sqrt{(n-2)/2}$, je $P\{\tau_i > \tau'\} = \dfrac{1}{n} P\{\tau_{(n)} > \tau'\}$, t. j. pravděpodobnosti, že τ_i překročí τ', je rovna pravděpodobnosti, že to bude právě i-té pozorování, které nabude největší hodnoty, násobené pravděpodobností, že největší τ bude větší než τ'. Tudíž $P\{\tau_{(n)} > \tau'\} = n P\{\tau_i > \tau'\}$, a pravděpodobnost na pravé straně se dá určit podle Thompsonových výsledků pomocí Studentova t- rozdělení. S užitím tohoto obratu mohli Pearson a Chandra Sekhar tabelovat kritické hodnoty $\tau_{(n)}(\alpha)$ pro některé rozsahy výběru a pro některé hladiny významnosti. Je-li tedy největší (nejmenší) hodnota ve výběru „podezřelá", počítáme charakteristiku $\tau_{(n)} = (x_{(n)} - \bar{x})/s$ $(\bar{x} - x_1)/s$ a zjistíme, je-li $\tau_{(n)} > \tau_{(n)}(\alpha)$. Pravděpodobnost nespravedlivého zamítnutí je α. Pearson a Chandra Sekhar se také zabývali účinností tohoto kriteria a zjistili, že je velmi účinné v případě, kdy ve výběru je jen jediná hodnota, zatížená hrubou chybou. Je-li více takových pozorování, účinnost kriteria klesá.

Obdobné kriterium, založené rovněž na odchylce extrémní hodnoty od aritmetického průměru, ale určené pro případ, kdy směrodatná odchylka základního rozdělení je známa, nebo kdy je k disposici její odhad, pořízený z jiného výběru, zpracoval r. 1948 K. R. Nair [13]. Tabeloval distribuční funkce veličin $\mu_n = \dfrac{x_{(n)} - \bar{x}}{\sigma}$, resp. $\dfrac{\bar{x} - x_{(1)}}{\sigma}$ a $v_{n,\nu} = \dfrac{x_{(n)} - x}{s_\nu} \left(\text{resp. } \dfrac{\bar{x} - x_{(1)}}{s_\nu}\right)$ a hodnoty $\mu_n(\alpha)$, pro které

$$P\left\{\frac{x_{(n)} - \bar{x}}{\sigma} > \mu_n(\alpha)\right\} = P\left\{\frac{\bar{x} - x_{(1)}}{\sigma} > \mu_n(\alpha)\right\} = \alpha$$

a $v_{n,\nu}(\alpha)$, pro které

$$P\{(x_{(n)} - \bar{x})/s_\nu > v_{n,\nu}(\alpha)\} = P\{(\bar{x} - x_{(1)})/s_\nu > v_{n,\nu}(\alpha)\} = \alpha^2).$$

Roku 1950 uveřejnil F. Grubbs práci [14], ve které rozšířil tabulku distribuční funkce kriteria μ_n, a podal odvození rozdělení dalších veličin:

---

2) s_ν je odhad směrodatné odchylky, pořízený z jiného výběru a založený na ν stupních volnosti.

$$F_n = S_n^2/S^2 = \sum_{1}^{n-1} (x_{(i)} - \bar{x}_1)^2 \Big/ \sum_{1}^{n} (x_{(i)} - \bar{x})^2, \quad \bar{x}_1 = \sum_{1}^{n-1} x_{(i)}/(n-1).$$

Kriterium F_n souvisí s dříve uvedeným τ vztahem: $F_n = 1 - \dfrac{1}{n-1}\,\tau^2$.

Velmi rychle se počítají kriteria Dixonova [15], založená na podílech rozpětí a rozdílů mezi sousedními hodnotami. Uvedeme jen jedno z nich:

$r = (x_{(n)} - x_{(n-1)})/(x_{(n)} - x_{(1)})$ pro test maximálního pozorování, resp.

$r = (x_{(2)} - x_{(1)})/(x_{(n)} - x_{(1)})$ pro test minimálního pozorování.

Dixon tabeloval hodnoty $r(\alpha)$, pro které

$$P\{r > r(\alpha)\} = \alpha.$$

Mimo to tabeloval i procentní body některých jiných podílů.

V následující tabulce jsou uvedeny pětiprocentní hodnoty některých z uvedených kriterií:

n	u_n (0,05)	$\tau_{(n)}$ (0,05)	r (0,05)
3	1,738	1,412	0,941
4	1,941	1,689	0,765
5	2,080	1,869	0,652
6	2,184	1,996	0,560
7	2,267	2,093	0,507
8	2,334	2,172	0,468
9	2,392	2,237	0,437
10	2,441	2,294	0,412
11	2,484	2,343	0,392
12	2,523	2,387	0,376
13	2,557	2,426	0,361
14	2,589	2,461	0,349
15	2,617	2,493	0,338
16	2,644	2,523	0,329
17	2,668	2,551	0,320
18	2,691	2,577	0,313
19	2,712	2,600	0,306
20	2,732	2,623	0,300

Na příklad pro 7 měření, při nichž při žádném nezapůsobil neobvyklý vliv (čili v náhodném výběru rozsahu 7 z normálního rozdělení) je

$$P\left\{\frac{x_{(7)} - \bar{x}}{\sigma} > 2{,}267\right\} = P\left\{\frac{\bar{x} - x_{(1)}}{\sigma} > 2{,}267\right\} =$$

$$= P\left\{\frac{x_{(7)} - \bar{x}}{s} > 2{,}093\right\} = P\left\{\frac{\bar{x} - x_{(1)}}{s} > 2{,}093\right\} =$$

$$= P\left\{\frac{x_{(7)} - x_{(6)}}{x_{(7)} - x_{(1)}} > 0{,}507\right\} = P\left\{\frac{x_{(2)} - x_{(1)}}{x_{(7)} - x_{(1)}} > 0{,}507\right\} = 0{,}05.$$

Ilustrujeme užití některých z uvedených kriterií na numerickém příkladě. V knížce „Měřické chyby a jejich vyrovnání" Dra B. Kladivo [16] jsou uvedeny výsledky 18 měření úhlu mezi dvěma směry. Výsledky, seřazeny podle velikosti, jsou:

83° 30′	30,25″	34,57	36,00
	33,16	34,75	36,14
	33,70	34,75	36,25
	33,75	34,77	36,50
	34,04	35,00	36,96
	34,25	35,25	37,50

$\bar{x} = 83° \, 30' \, 34{,}866''$, $s = 1{,}62''$.

Otestujeme významnost odchylky nejmenšího pozorování od ostatních. $(x - x_{(1)})/s = (4{,}866 - 0{,}25) : 1{,}62 = 2{,}849$. Z tabulky zjistíme, že při výběru rozsahu 18 je pětiprocentní hodnota kriteria $(\bar{x} - x_{(1)})/s$ rovna 2,577. Pozorovaná hodnota je větší, závěr by tedy zněl, že nejmenší pozorování je zatíženou hrubou chybou, a při dalším zpracování by se použilo jen zbývajících 17 měření. Pravděpodobnost, že v řadě měření bez hrubých chyb, to jest, pořízených za stejných podmínek, nabude veličina $(\bar{x} - x_{(1)})/s$ hodnoty tak vysoké, jako v uvedené řadě, je totiž menší než 0,05.

Kriterium r nabývá v uvedené řadě hodnoty

$$(x_{(2)} - x_{(1)})/(x_{(18)} - x_{(1)}) = 2{,}91/7{,}25 = 0{,}401.$$

Pětiprocentní hodnota kriteria r, nalezená ve třetím sloupci tabulky, je 0,313. Pozorovaná hodnota 0,401 je větší. Tedy kriterium r by také vedlo k závěru, že nejmenší hodnota je zatížena hrubou chybou a měla by se při zpracování vynechat.

Nyní zbývá zmínit se o pokusech o vyšetření účinnosti uvedených kriterií při odhalování nesprávnosti testované hypothesy. Většinou je účinnost statistického testu vyjádřena silofunkcí, která udává pravděpodobnost zamítnutí testované hypothesy, když je správná některá alternativní hypothesa. V našem případě tedy silofunkce udává pravděpodobnost, že kriterium nabude významné hodnoty, když v řadě měření skutečně je krajní hodnota zatížena hrubou chybou. Tato pravděpodobnost by se však počítala velmi obtížně. Dosud byla uveřejněna jediná práce, ve které jsou srovnány silofunkce jednotlivých uvedených kriterií [17]. Její autor, W. J. Dixon, zjišťoval (právě pro velkou obtížnost numerických výpočtů) silofunkce empiricky. Prováděl to tak, že sestrojil model normálního rozdělení s parametry μ, σ, uměle pořídil výběr rozsahu $n - 1$ a přidal jedno pozorování ze základního souboru normálního s parametry $\mu + d\sigma$, σ. Tak dostal výběr, ve kterém $n - 1$ pozorování pocházelo ze základního souboru $N(\mu, \sigma)$ a jedno ze základního souboru $N(\mu + d\sigma, \sigma)$. Když to jedno pozorování bylo extremní hodnotou, provedl test významnosti jeho odchylky od ostatních. Podíl počtu případů, kdy test dal významný výsledek, k celkovému počtu výběrů dal odhad síly testu. To provedl Dixon pro většinu uvedených kriterií, pro dva výběrové rozsahy a pro několik hodnot d. Ukázalo se, že nejúčinnější jsou kriteria typu $\dfrac{x_{(n)} - \bar{x}}{\sigma}$, a $\dfrac{x_n - x}{s_\nu}$.

Pro aplikace na chyby měření není silofunkce tím nejvhodnějším ukazatelem účinnosti kriteria hrubých chyb. Při zpracování řady měření jde totiž o odhad určité veličiny, a ne o rozhodnutí, zda jedno pozorování je z jiného souboru. Odhad má být zbaven vlivu hrubých chyb měření. Je tedy třeba vyhledat pokud možno všechny hrubé chyby a odpovídající pozorování odstranit. Proto se při vyřazování hrubých chyb kriterium aplikuje několikrát po sobě. Vede-li test k vyloučení nejvzdálenější hodnoty, vypočte se ze zbývajících $\bar{x}$ a s a testuje se znovu, atd., až dostaneme řadu, jež vyhovuje zvolenému kriteriu.

Provede se tedy jakýsi vylučovací postup, který patří k proceduře odhadu. Jeho součástí je konstanta α, hladina významnosti, na které se jednotlivé kroky tohoto postupu uskutečňují. Chceme-li studovat, jakého zlepšení odhadu se takovým vylučovacím postupem dosáhne, musíme stanovit nejprve nějaký model hrubých chyb. Na př. při měření veličiny μ zasáhne s pravděpodobností γ nějaký rušivý vliv, posunující hodnotu o $k\,\sigma$ do prava. Jinými slovy: provádíme výběr z normálního rozdělení. Rozdělení, z kterého se prvek vybere, je samo určeno náhodně. S pravděpodobností $1-\gamma$ má parametry μ, σ, s pravděpodobností γ parametry $\mu + k\,\sigma$, σ. Tento model odpovídá na př. této situaci: Pracovník se při řadě odečítání na stupnici přístroje zmýlí průměrně ve $100\,\gamma$ % případů o celý velký dílek stupnice, který odpovídá k-násobku směrodatné odchylky měření. Při takové možnosti hrubých chyb je odhad veličiny μ pomocí obyčejného aritmetického průměru jednostranný, $Ex = \mu + \gamma\,k\,\sigma$. Jinak můžeme chybu odhadu měřit střední hodnotou čtverce odchylky od skutečné hodnoty, lomenou rozptylem odhadu v nepřítomnosti rušivých vlivů, $E(\bar{x} - \mu)^2/\sigma^2 = Q$.

Nyní si představme, že před výpočtem $\bar{x}$ aplikujeme některé kriterium hrubých chyb. Tím vyřadíme některé hrubé chyby, možná všechny, možná také některá „dobrá" pozorování. Tím změníme Q. Bylo by nejlepší zvolit takový vylučovací postup, který by minimalisoval Q. Zatím k nalezení takového postupu neukázal nikdo jinou cestu než empirickou: uměle pořizovat výběry z různých modelů, odpovídajících různým experimentálním podmínkám, vylučovat extrémy pomocí různých kriterií a empiricky stanovit zvolenou míru chyby odhadu. Pro některé hodnoty γ a k to provedl Dixon (1953) se svými kriterii r a dospěl k jistým pravidlům pro volbu hladiny α v závislosti na γ a k [18]. Pro některé kombinace hodnot γ a k se ukázalo výhodnějším užívat jako odhadu veličiny μ výběrového mediánu, který je méně ovlivněn extrémními hodnotami než aritmetický průměr. Dixonova pravidla tedy stanoví volbu α v závislosti na hodnotách γ a k. To znamená, že jich lze s úspěchem užít jen tam, kde je přibližně známo, v jakém směru a jak často rušivé vlivy působí. To je jen projev známé zkušenosti, že znalost alternativ, proti kterým má být statistický rozhodovací postup zaměřen, umožňuje volbu vydatnějších method.

Literatura

[1] *Criterion for the Rejection of Doubtful Observations*, Astr. J., sv. II, 1952.
[2] Stewart, *Peirce's Criterion*, Popular Astronomy, sv. XXVIII, 1920.
[3] W. Chauvenet, *A Manual of Spherical and Practical Astronomy*, díl II, vyd. 5, 1876.
[4] I. M. Malikov, *Osnovy Metrologii*, vyd. 2., Moskva 1949, kap. 5, par. 45.
[5] J. O. Irwin, *On a Criterion for the Rejection of Outlaying Observations*, Biometrika 17, 1925.
[6] J. O. Irwin, *The Further Theory of Francis Galton's Individual Difference Problem*, tamtéž.
[7] L. H. C. Tippett, *On the Extreme Individuals and the Range of Samples Taken from a Normal Distribution*, tamtéž.
[8] E. S. Pearson, *Further Note on the Distribution of Range*, Biometrika 18, 1926, str. 173.
[9] W. Gosset („Student"), *Errors of Routine Analysis*, Biometrika 19, 1927, str. 151.
[10] Mc Kay, *The Distribution of the Difference between the Extreme Observation and the Sample Mean in Samples of N from a Normal Universe*, Biometrika 27, 1935, str. 466.
[11] W. R. Thompson, *On a Criterion for the Rejection of Observations and the Distribution of the Ration of Deviation to Sample Standard Deviation*, Ann. of Math. Statistics, 6, 1935, str. 214.
[12] E. S. Pearson a. Chandra Sekhar, *The Efficiency of Statistical Tools and a Criterion for Testing Outlaying Observations*, Biometrika 28, 1936, str. 308.
[13] K. R. Nair, *The Distribution of the Extreme Deviate and its Studentized Form*, Biometrika 35, 1948, str. 118.
[14] F. Grubbs, *Sample Criteria for Testing Outlaying Observations*, Ann. of Math. Statistics, 21, 1950, str. 27.
[15] W. J. Dixon, *Rations Involving Extreme Values*, tamtéž, 22, 1951, str. 68.
[16] Dr B. Kladivo, *Měřické chyby a jejich vyrovnávání*, JČMF.
[17] W. J. Dixon, *Analysis of Extreme Values*, Ann. of Math. Statistics, 21, 1950, str. 488.
[18] W. J. Dixon, *Processing Data for Outlayers*, Biometrics 9, 1953, str. 74.