

Hodnocení analýz tuhých paliv

Kubant J., Lustigová M.

Praha 2014

ANOTACE

Předložená práce směřuje k prezentaci upravené nové metody správného hodnocení a přepočtů chemických a fyzikálních analýz uhlí pomocí lineárních regresních závislostí znaků jakosti na obsahu popela, resp. reciproké hodnotě obsahu popela. Převážná část sdělení je věnována přehledným postupům zpracování dat formou jednoduchých algoritmů umožňujících rychlou orientaci v dané problematice. Způsob hodnocení uvedený ve třetí kapitole je podkladem pro revizi ČSN 44 1364.

OBSAH

1. ÚVOD	6
2. VYBRANÉ ELEMENTÁRNÍ POJMY MATEMATICKÉ STATISTIKY	8
2.1 Chyby, přesnost a správnost, kumulace chyb.....	8
2.2 Rozsah souboru	10
2.2.1 Normalizovaný způsob nejčastěji dvou paralelních stanovení	11
2.2.2 Malý rozsah výběru.....	11
2.2.3 Velký rozsah výběru	12
2.3 Vylučování odlehlých výsledků jednorozměrných dat.....	14
2.3.1 Dean - Dixonův test	15
2.3.2 Grubbsův test	15
2.3.3 Posouzení odlehlosti.....	15
2.4 Test významnosti rozdílů $x - \mu_0$	16
2.5 Vyhledávání soustavných chyb – odchylek.....	16
2.5.1 Párový t-test	17
2.5.2 Znaménkový test	17
2.6 Střední hodnoty, charakteristiky úrovně (polohy).....	18
2.6.1 Medián x	18
2.6.2 Aritmetický průměr x	18
2.6.3 Vážený aritmetický průměr x_v	18
2.6.4 Harmonický průměr x_h	19
2.6.5 Vážený harmonický průměr x_{hv}	19
2.6.6 Geometrický průměr x_g	19
2.6.7 Vážený geometrický průměr x_{gv}	19
2.6.8 Kvadratický průměr x_k	20
2.6.9 Vážený kvadratický průměr x_{kv}	20
2.6.10 Modus x	20

2.7 Charakteristiky variability	21
2.7.1 Variační rozpětí R	21
2.7.2 Průměrná odchylka $ d $	21
2.7.3 Odhad rozptylu základního souboru σ^2	22
2.7.4 Odhad disperze $D(s^2)$	22
2.7.5 Dvoustranný konfidenční interval rozptylu σ^2	22
2.7.6 Výběrová směrodatná odchylka s	22
2.7.7 Odhad směrodatné odchylky základního souboru σ	23
2.7.8 Odhad disperze $D(s.k_v)$	23
2.7.9 Odhad směrodatné odchylky aritmetického průměru s_x	23
2.7.10 Relativní směrodatná odchylka s_{rel}	23
2.7.11 Interval spolehlivosti pro parametr σ	23
2.7.12 Hromadění chyb při základních operacích.....	24
2.7.13 Rozptyl statistického souboru S^2	24
2.7.14 Vážený rozptyl statistického souboru S_v^2	24
2.7.15 Celkový vážený rozptyl z rozptylů skupinových souborů.....	24
2.8 Ověřování normality rozdělení četností	25
2.8.1 Charakteristika šikmosti.....	25
2.8.2 Charakteristika špičatosti	26
2.9 Lineární regresní závislosti.....	26
2.9.1 Regresní přímka	26
2.9.1.1 Korelační koeficient r_{yx}	28
2.9.1.2 Regresní přímka $Y = a_1 + b_1 \cdot x$ podle varianty $\sigma_x = 0, \sigma_y \neq 0$	28
2.9.1.3 Regresní přímka $Y = a_2 + b_2 \cdot x$ podle varianty $\sigma_x \neq 0, \sigma_y = 0$	28
2.9.1.4 Regresní přímka $Y = a_3 + b_3 \cdot x$ podle varianty $\sigma_x \neq 0, \sigma_y \neq 0$	29
2.9.1.5 Intervaly spolehlivosti regresní přímky.....	29
2.9.1.6 Praktické využití regresní přímky.....	30
2.9.1.6.1 Youdenova metoda identifikace druhů chyb (odchylek)	30
2.9.1.6.2 Tři varianty odhadu empirických regresních přímek	32
2.9.1.7 Vylučování bodů odlehlých od regresní přímky.....	34

2.9.2. Regresní závislost se dvěma nezávisle proměnnými – regresní rovina.	35
2.9.2.1. Korelační koeficienty	36
2.9.2.2. Koeficienty β	36
2.9.2.3. Konstanty regresní rovnice roviny $Y = a_{y.12} + b_{y1.2} \cdot x_1 + b_{y2.1} \cdot x_2$	37
2.9.2.4. Směrodatné odchylky parametrů regresní roviny	37
2.10 Seznam tabulek I až VII	39
Tabulka I: Kritické hodnoty t_α rozdělení	41
Tabulky IIa,b: Kvantily u_p normálního rozdělení	42
Tabulka III: Kritické hodnoty Dean-Dixonova testu	43
Tabulka IV: Kritické hodnoty Grubbsova testu	44
Tabulka V: Kritické hodnoty znaménkového testu	45
Tabulka VI: Kvantily rozdělení $\chi^2(v)$	47
Tabulka VII: Kritické korelační koeficienty r_α	49
Literatura 2. kapitoly	50
3. HODNOCENÍ ANALÝZ TUHÝCH PALIV	51
3.1 Fosilní tuhá paliva	51
3.1.1 Dosavadní přepočty na hořlavinu	51
3.1.2 Teoretické principy pracovní hypotézy	52
3.1.3 Odhady hodnot základních charakteristik výběru	54
3.1.4 Vlastnosti námi formulovaných regresních závislostí	57
3.1.5 Odběr vzorků	59
3.1.6 Požadované vlastnosti výběru vzorků uhlí	60
3.1.7 Využití výsledků hodnocení.	60
3.1.8 Postup při výpočtech	61
Literatura 3. kapitoly	62

1. ÚVOD

Tuhými palivy z recentních a fosilních zdrojů rozumíme hmoty určené k získání tepelné energie přímým spalováním nebo po zušlechtní, případně určené k prvotní přeměně na kapalná či plynná paliva. Sem se řadí biomasa, tj. dřevo, byliny, části rostlin nesoucí semena a kaustobiolity, tj. rašelina, lignit, hnědé a černé uhlí a antracit.

Před využitím lze jakost tuhých paliv zlepšit úpravou obsahu balastu (vody, popelovin), úpravou do tvarů vhodné velikosti (např. briketováním, peletizací), případně tepelným zpracováním (dřevěné uhlí, polokoks, koks). Všechny naznačené způsoby úprav a využití tuhých paliv vyžadují více či méně náročné technologické zpracování. Jak průzkum ložisek uhlí, tak projektování a provoz technologií se neobejde bez znalosti jakosti konkrétního paliva potřebného rozsahu. K tomu slouží analýzy vhodně odebraných a upravených reprezentativních vzorků tuhých paliv, potřebné správnosti a předepsané opakovatelnosti, vyjádřené hodnotami znaků jakosti v původním, bezvodém stavu a v hořlavině – organické hmotě.

Tuhá paliva lze chápat jako ternární směs vody, popelovin a organické hmoty. Zjednodušeně a nepřesně lze sumu těchto tří složek vyjádřit rovnicí $100\% = \text{voda} + \text{popel} + \text{hořlavina}$. Obsah popela po spálení paliva zpravidla neodpovídá obsahu popelovin před spálením. Chyba při běžném používání zmíněné nepřesné rovnice, při přepočtu výsledků analýz na „hořlavinu“, je tím větší, čím větší je obsah popelovin, resp. popela.

Recentní biomasa obsahuje většinou pod 5% popela, výjimečně maximálně kolem 10% u slámy, což vylučuje možnost hodnocení jejich rozborů pomocí regresních vztahů mezi znaky jakosti organické hmoty biopaliv a obsahem popela. U biopaliv lze v oblasti jejich analýzy konstatovat, že v rámci harmonizace norem ČSN s EU byly v ČR normy pro biopaliva prakticky přijaty formou překladů bez kritického odborného hlediska. Tím došlo i k převzetí analytického zrnění vzorků pod 1mm, které negativně ovlivňuje opakovatelnost a reprodukovatelnost výsledků analýz, a to jak v důsledku zvýšení heterogenity vzorků, tak při nedokonalém spalování (stanovení obsahu popela a spalného tepla), ve srovnání s analytickými vzorky uhlí zrnění pod 0,2mm. Pokud akceptace zrnění pod 1mm měla ulehčit úpravu vzorků biopaliv, lze konstatovat, že největší problémy s úpravou vzorků biopaliv jsou právě v počáteční fázi úpravy. Zmenšení zrna z 1mm na 0,2mm je velmi snadné ve vibračním mlýnu s náplní válec-mezikruží. Z uvedených důvodů je potřebné sjednotit úpravu analytických vzorků všech tuhých paliv na zrna pod 0,2mm. V řadě případů jsou spotřebitelé využívající společně tuhých fosilních i biopaliv vybaveni úpravami vzorků na zrna pod 0,2mm.

U vzorků tuhých fosilních paliv lze očekávat asi 5–60% popela a při průzkumu sloje i vzorky s téměř 100% popelovin. V takových případech přepočty na „hořlavinu“ nabývají absurdních hodnot. Existuje jediná možná výjimka, a to tehdy, je-li u vzorku hodnota popelového faktoru (f_A) rovna přesně 1. Tehdy se obsah popelovin (M) rovná obsahu popela (A), tudíž $M/A = f_A = 1$. Zcela obecně lze konstatovat, že popelový faktor není konstantou, ale je závislý na obsahu popela a jeho složení. V předložené publikaci je proto věnována mimořádná pozornost z teoretického i praktického hlediska řešení problému rozdílu mezi obsahem popela a popelovin a s tím souvisejícímu odhadu popelového faktoru, obsahu hydrátové vody, regresním vztahům mezi znaky jakosti, vázaným a volným popelovinám a dalším charakteristikám. Jak se ukázalo v souvislosti s řešením této problematiky, lze odhadnout i řadu charakteristik kaustobiolitů uhlénné řady dosud nestanovitelných.

U řady postupů, které jsou uzančnými metodami, je nutné respektovat přesně normou předepsané podmínky a postupy (teplota, čas, výška vrstvy, navážka atd.), aby byly získány

výsledky reprodukovatelné v různých laboratořích. U hodnot znaků jakosti, které byly vyjádřeny v původních stavech, musí být uvedeny údaje o obsahu vody a popela, podobně u znaků vyjádřených v bezvodých stavech musí být uvedeny obsahy popela vždy v přesně definovaných stavech. Podobně u všech fyzikálních a fyzikálněchemických vlastností musí být vždy uvedeny údaje o obsazích vody a popela, které byly v palivu obsaženy při jejich měření a analýze, neboť bez nich neposkytnou informaci s dostatečnou vypovídací hodnotou. Vzorky fosilních paliv jsou na vzduchu nestálé, a proto musí být věnována pozornost i vlivu času a podmínek uložení vzorků od odběru k analýze. Je zřejmé, že chemické složení paliva ve sloji se bude lišit od téhož paliva ve mlýnu před kotlem po pobytu na skládce.

Vznik této publikace má několik podnětů. Jednak je to již přes 40 let od posledního vydání knihy v českém jazyce o analýze paliv [1]. Je to i absence podrobnějšího výkladu o validitě hodnot znaků jakosti přepočtených na „hořlavinu“, včetně nové metody jednoduchého odhadu popelového faktoru a dalších charakteristik hodnoceného výběru. Dále to je snaha o rozšíření užívání jednoduchých lineárních regresních vztahů mezi znaky jakosti uhlí, jako metody hodnocení výsledků analýz do palivářských laboratoří, včetně prezentace elementárních metod matematické statistiky formou jednoduchých algoritmů. Velká pozornost věnovaná uhlí v předložené práci je motivována tím, že stále podstatná část energetických potřeb je kryta využitím uhlí.

Závěrem bychom chtěli připomenout dlouholetou tradici analýzy paliv v Čechách, a to již od doby rakouské monarchie, např. [2,3]. Zvláště je pak třeba upozornit na práci SCHULZOVU [4], jako první podklad pro normalizaci zkušebních metod tuhých paliv v ČSR, práci HUBÁČKOVU [5], obsáhle se zabývající složením a klasifikací československých tuhých paliv, dále ŠIMKOVU a LUDMILOVU knihu [6] pojednávající podrobně o rozbořech tuhých paliv a konečně knihu o chemii uhlí HUBÁČKA a kol. [7] shrnující poznatky o uhlí v tehdejší době. Zmíněná česká odborná literatura pojednávající o složení a analýze tuhých paliv, včetně publikací z oborů geochemie, geologie, petrografie apod., však bohužel neřeší přijatelným způsobem problém rozdílu mezi obsahem popelovin a popela. Prvním pokusem o řešení problému metodou matematické statistiky byla publikace [8], která však problém nedořešila. Přesto byly principy metody převzaty do ČSN 44 1364 [9]. Předložená studie je mimo dalších aspektů hlavně pokusem o dořešení problému výpočtů správných výsledků analýz tuhých paliv.

Literatura 1. kapitoly

- [1] ŠEŠULKA, V.: Analýza paliv. SNTL Praha, 1970, str. 13 až 92.
- [2] VYSOKÝ, A.: Kokovací zkouška pro hnědé uhlí. Zpráva ústavu pro techn. paliv. 1905, str. 58.
- [3] SCHULZ, F.: Určování síry v palivech. Chem. Listy V. 1911.
- [4] SCHULZ, F.: Přehled method pro chemický rozbor a zkoušení uhlí a koku. Ústav M.A.P., 1931.
- [5] HUBÁČEK, J.: Tuhá paliva československé republiky. Matice hornicko-hutnická, 1948, 340 s.
- [6] ŠIMEK, B. G. – LUDMILA, J.: Rozbory tuhých paliv. T. – V. Vydavatelství Praha, 1951.
- [7] HUBÁČEK, J. – a kol.: Chemie uhlí. SNTL Praha, 1962.
- [8] KUBANT, J.: Uhlí, 34, 1986, č.10, s. 406-413.
- [9] ČSN 44 1364: VÝPOČET OBSAHU POPELOVIN, HYDRÁTOVÉ VODY A HODNOT ZNAKŮ JAKOSTI V ORGANICKÉ HMOTĚ. 9.12.1986.

2. VYBRANÉ ELEMENTÁRNÍ POJMY MATEMATICKÉ STATISTIKY

Jednotlivé hodnoty znaků jakosti implicitně obsahují víc informací, než se na první pohled zdá. Je-li dostatečný počet pozorování a jsou-li vybrána vhodná a dostatečně přesná data, může použití vhodných metod a správná interpretace výsledků výpočtů poskytnout více informací než samotná jednotlivá data, která jsou nositeli těchto informací. U regresních závislostí je nutné předem prověřit vhodnost použité funkce v souladu s teorií a empirickými zkušenostmi a najít obecnou, ale zjednodušenou abstrakci daného problému. Aproximační funkce má vykazovat co nejmenší odchylky od experimentálních hodnot. Proto je také nutné posoudit, zda výběrový soubor neobsahuje hodnoty zatížené extrémními chybami, které je nutné z této množiny hodnocených dat vyloučit, pokud do ní svými vlastnostmi nepatří.

2.1 Chyby, přesnost a správnost, kumulace chyb

V naší problematice bývá rozptýl znaků jakosti vzorků téhož původu dán především variabilitou obsahu vody a popelovin, dále variabilitou složení organické hmoty a popelovin vzorků dané lokality. Konečně se zde uplatňují vlivy chyb od odběrů vzorků a jejich úpravy až po chyby výpočtů analýz. Někdy je nutné uvažovat i o vlivu složení analyzovaných vzorků na rozptýl výsledků získaných uzančnickými metodami vypracovanými případně pro odlišné typy chemického složení a vlastností vzorků. Významný může být i vliv oxidace od odběru po analýzu vzorků.

Výsledky i opakovaných stanovení znaku jakosti téhož vzorku se mohou lišit od skutečné hodnoty μ , která je neznámá a tato diference je hodnocena jako chyba skutečná (úplná)

$$\varepsilon_i = x_i - \mu \quad (2.1)$$

kde: ε_i = skutečná úplná chyba i-tého stanovení znaku, zahrnuje náhodnou i soustavnou složku,

x_i = hodnota i-tého stanovení znaku,

μ = skutečná, správná hodnota znaku.

Použijeme-li místo neznámé hodnoty μ aritmetický průměr $\bar{x}$, vypočtený i z rozsáhlého souboru opakovaných stanovení, nejedná-li se o soubor základní, je diference hodnocena jako chyba zdánlivá

$$e_i = x_i - \bar{x} \quad (2.2)$$

kde: e_i = zdánlivá chyba i-tého stanovení.

Do úplné chyby ε_i se zahrnují chyby, které lze zhruba rozdělit na chyby náhodné, chyby soustavné a chyby hrubé či omyly. Mezi nimi je v podstatě spojitý přechod.

Náhodné chyby (d_1) mohou nabývat různé velikosti a znaménka; jejich hodnoty kolísají kolem nuly, přičemž střední hodnota, ze všech možných hodnot, je rovna nule. Náhodné chyby jsou vzájemně nezávislé a nepředvídatelné. Každá z nich může vzniknout kombinací většího počtu náhodných chyb.

Soustavné chyby (d_2), kterých je mnoho druhů [1], buď nadhodnocují, nebo podhodnocují stanovené hodnoty. Kolísají kolem nenulové hodnoty a obvykle zhoršují správnost výsledků víc než chyby náhodné. Mohou být vyvolány chybným vzorkováním, vadnými váhami, chybným měřením objemu, teploty, napětí, nastavením vlnové délky apod. Při absenci hrubých chyb a rozsahu $n \rightarrow \infty$ je

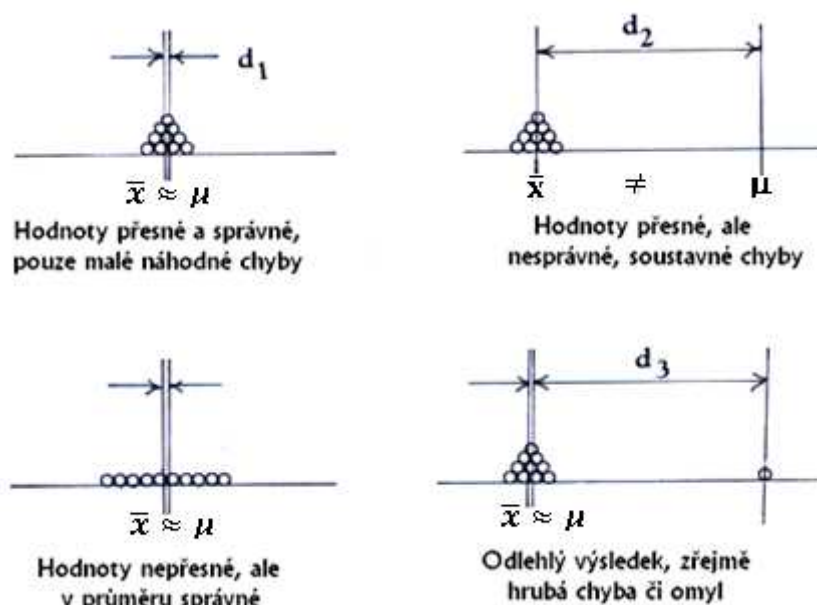
$$(\sum_{i=1}^n \varepsilon_i) : n = \overline{d_2}, \text{ protože v takovém případě } (\sum_{i=1}^n d_{1i}) = 0 \quad (2.3)$$

kde: $\overline{d_2}$ = střední hodnota soustavných chyb, výsledek vzájemných kombinací.

Chyby hrubé a omyly (d_3) nejsou způsobeny objektivními podmínkami stanovení znaku, ale nesprávnými úkony pracovníka, případně použitím nevhodného postupu, ztrátou části analytu, nečistotami v chemikáliích, chybným zápisem apod. Často na sebe hrubá chyba upozorní vybočením z obvyklých hodnot a vyžádá si tak prověření postupu stanovení, případně vyloučení odlehlé hodnoty a opakování jejího stanovení za zvýšené pozornosti k průběhu analýzy.

Kumulace chyb. Chyby, kterými jsou zatíženy konečné hodnoty znaků jakosti, jsou výsledkem kumulace chyb vznikajících při řadě úkonů souvisejících s procesem odběru vzorků, jejich úpravy a zmenšování hmotnosti a objemu a konečně i analýzy vzorků. Každý z těchto úkonů může být nezávislým zdrojem různých chyb různého znaménka a velikosti. Při odběrech vzorků, které bývají obvykle největším zdrojem chyb, je rozhodující heterogenita odebíraného vzorku a reprezentativnost odebraného vzorku. Při redukci objemu (hmotnosti) vzorku dělením je rozhodující velikost zrna. Čím menší je rozměr zrna děleného vzorku, tím menší je chyba dělení homogenizovaného vzorku. Podrobnou analýzou chyb se zabývá BÖHM [1], vzorkováním tuhých paliv RESLER, RIEDL [2], obecně vzorkováním ČSN 01 5110 a 01 5111 [3,4].

V souvislosti se zmíněnými druhy chyb je třeba vyjasnit i pojmy přesnost a správnost. Velmi názorně tyto pojmy osvětlil graficky ECKSCHLAGER [5]. Proto jsme použili podobného způsobu, viz obr. 2.1.



Obr.2.1: Grafické znázornění přesnosti a správnosti a odlehlosti výsledků

Pro rozdíly mezi paralelními výsledky v normalizovaných metodách stanovení a měření jsou předepsány povolené tolerance, které jsou definovány jako opakovatelnost a reprodukovatelnost.

Opakovatelnost představující těsnost shody mezi navzájem nezávislými výsledky zkoušek, opakovaným použitím téže zkušební metody, na identickém materiálu, v téže laboratoři, týmem pracovníkem, za použití týchž přístrojů a zařízení, během krátkého časového rozmezí.

Reprodukovatelnost představující povolenou těsnost shody mezi laboratořemi, při použití téže zkušební metody, na identickém materiálu, různými pracovníky, za použití různých přístrojů a zařízení.

Zmíněné povolené těsnosti shody výsledků jsou součinem ($s \cdot f \cdot \sqrt{2}$), tj. příslušných směrodatných odchylek získaných při mezilaboratorních zkouškách, kde činitel $f \approx 2$ odpovídá pravděpodobnosti 95%, za předpokladu dostatečně rozsáhlého souboru, s přibližně normálním rozdělením chyb a $\sqrt{2}$ je uplatněna s ohledem na to, že se jedná o rozdíl mezi dvěma hodnotami.

2.2 Rozsah souboru

Základním souborem se rozumí množina všech možných jednotek se společnou vlastností- znakem. Znak je sledovaná kvantitativní či kvalitativní vlastnost, kterou lze vyjádřit číslem a měřicí jednotkou. Výběrový soubor (zkrácený termín „výběr“) je pak konečná podmnožina základního souboru. Předmět (vzorek), na němž lze provést pozorování nebo výsledek pozorování, je jednotka (prvek) pozorování (zkrácený termín „jednotka“). Rozsahem souboru je počet jednotek (prvků) v souboru značený n . Výpočet charakteristik (parametrů) výběru slouží též k odhadu charakteristik (parametrů) základního souboru, které neznáme. Rozsah základního souboru v naší problematice obvykle velmi rozsáhlý, až prakticky nekonečný, nám neumožňuje jeho zpracování v celém rozsahu. Je přirozené, že čím větší je rozsah výběru, tím menší je vliv kompenzujících se náhodných chyb na výsledek výpočtu – odhadu průměrné reprezentativní hodnoty. Podmínkou k vyhovující přesnosti a správnosti odhadů parametrů základního souboru je i absence soustavných a hrubých chyb. V praxi jsme často postaveni před rozhodnutí, jaký rozsah výběru použít. Malé rozsahy poskytují odhady méně přesné, velké rozsahy výběrů jsou pro zpracování časově náročné a nákladné. V běžné literatuře doporučované vztahy pro odhad rozsahu výběru – např. při neznámých hodnotách μ a σ

$$n = \left(\frac{t_{1-\alpha/2}^2}{\Delta_{max}^2} \right) s^2 \quad (2.4)$$

kde: n = rozsah výběru (počet jednotek),

t = kvantil „Studentova“ rozdělení, (tab.I. s. 41)

Δ_{max} = zvolená maximální absolutní chyba, která nemá být překročena; zde představuje poloviční rozpětí dvoustranného konfidenčního intervalu,

s = výběrová směrodatná odchylka viz (2.38a,b)

σ = směrodatná odchylka základního souboru,

α = hladina významnosti,

a další vztahy, např. [6] na bázi relativních chyb, však bohužel postrádají informaci o limitním rozsahu n , který nemá být podkročen a dále jednoduchou, přibližnou informaci o efektivnosti zvětšování rozsahu n . Proto byly v této práci předloženy přibližné informace o minimálních a přijatelných rozsazích n , které je možno použít, není-li použítí maximální chyby např. podle vztahu (2.4) nevyhnutelnou podmínkou, a tím i daný rozsah n . Zajímavý, ale výpočetně poněkud náročnější odhad minimálního výběrového rozsahu uvádějí MELOUN a MILITKÝ [17], který vychází ze stanovené špičatosti výběrového rozdělení a je vztažen k relativní chybě směrodatné odchylky (obvykle 10%); uveden je i příklad výpočtů pro pět různých druhů rozdělení. Problematika určování a hodnocení rozsahu výběru, kterou se zabýváme vždy jako prvou, lze rozdělit do několika úrovní.

2.2.1 Normalizovaný způsob nejčastěji dvou paralelních stanovení

Povolené tolerance dané dříve zmíněnou opakovatelností a reprodukovatelností v normách a podobných předpisech přesně vymezují přijatelnou shodnost výsledků stanovení znaku jakosti na základě dříve zjištěné variability zvolené metodiky stanovení znaku jakosti. V některých případech je přijatelný i průměr tří stanovení, pokud první a druhé stanovení překračuje nevýznamně povolenou toleranci ($|\Delta_1| = x_1 - x_2$) a třetí stanovení vykazuje takovou hodnotu, že $x_1 < x_3 < x_2$, kdy $|\Delta_1| = x_3 - x_1$ a $|\Delta_2| = x_3 - x_2$ vyhovují povolené toleranci, případně průměr z nejbližších dvou hodnot. Jiné způsoby požadující zpracování rozsáhlejšího souboru, např. při hodnocení hustoty a mechanické pevnosti pelet a briket, jsou dány příslušnými normativními předpisy.

2.2.2 Malý rozsah výběru

Přibližný odhad minimálního rozsahu n s využitím hodnot kvantilů „Studentova“ t rozdělení je vhodný tam, kde při minimálním počtu opakovaných stanovení je snaha získat přijatelnou přesnost průměrných výsledků $\bar{x}$, při zvolené hladině významnosti α , chybí-li informace o parametrech základního souboru. Hodnotí se buď náhodné výběry analýz sypkých homogenních vzorků (zrnění $< 0,2\text{mm}$, případně menší), nebo jiných vzorků po odstranění makroheterogenity. Variabilita opakovaných výsledků stanovení v těchto případech je daná spíše přesností metodiky než vlivem heterogenity vzorku. Minimální rozsahy výběrů bývají motivovány nákladností a časovou náročností analýz. Smyslem dále uvedeného postupu je odhad minimálního rozsahu výběru, který nemá být podkročen.

Při úvaze o přesnosti odhadu parametru μ pomocí výběru velmi malého rozsahu lze šíři dvoustranného konfidenčního intervalu spolehlivosti $I_{\mu t}$ pokrývajícího parametr μ formulovat součinem

$$I_{\mu t} = \left(\frac{2 \cdot t_{\alpha}(v)}{\sqrt{n}} \right) \cdot s = q_t \cdot s \quad (2.5)$$

kde součinitel q_t je funkcí rozsahu výběru n , dále zvolené pravděpodobnosti $P=(1-\alpha) \cdot 100\%$ a jí odpovídajících kvantilů rozdělení t pro dvoustranný kritický obor, což lze snadno odvodit z obvyklé formulace oboustranného konfidenčního intervalu, tj. neostře nerovnosti

$$\bar{x} - \left(\frac{t_{\alpha}(v)}{\sqrt{n}} \right) \cdot s \leq \mu \leq \bar{x} + \left(\frac{t_{\alpha}(v)}{\sqrt{n}} \right) \cdot s \quad (2.6)$$

kde: t_{α} = kvantil rozdělení t , pro dvoustranný kritický obor, viz tab. I., str. 41

α = hladina významnosti,

(v) = počet stupňů volnosti, v tomto případě $(v) = (n-1)$

s = výběrová směrodatná odchylka daná vlastnostmi výběru, viz (2.38a,b).

K odhadu intervalu spolehlivosti $I_{\mu t}$ se nedoporučuje soubor s hodnotou $(v) < 4$ [1]. Na obr.

2.2 jsou graficky znázorněny funkční závislosti $q_t = f(n)$ pro různé pravděpodobnosti $P\%$. Z jejich průběhů je zřejmé, že se hodnoty q_t se zvětšujícími se rozsahy výběrů n zmenšují, zprvu velmi účinně až do dosažení vrcholu křivek, kde poloměr křivosti dosahuje minima ($n_{\min}$). Za vrcholem se zmenšování hodnot q_t rychle zpomaluje, až nakonec křivky asymptoticky konvergují k 0 při zvolené pravděpodobnosti a $n \rightarrow \infty$. Z průběhů závislostí lze usuzovat, že minimalizace rozsahu výběru n by měla být limitována hodnotou označenou n_v (v = vrchol). Důvod limitace n je jasný; při podkročení rozsahu $n_{\min}$ se rychle zvětšuje hodnota q_t , čímž se rozšiřují meze konfidenčního intervalu a enormně vzrůstá nepřesnost odhadů parametru μ . Informace k velmi malému výběru

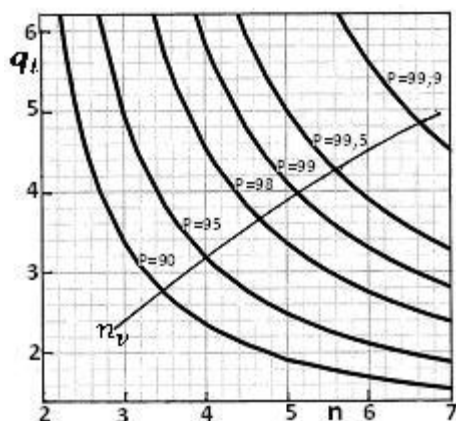
viz tabulka 2.1. Protože s není konstantou a výběr je velmi malý a $n_{\min}$ se zaokrouhluje, je odhad $I_{\mu t}$ podle rovnice (2.7) pouze přibližný.

Tabulka 2.1 Vlastnosti velmi malých výběrů

P%	u_p pro odhad $n_{\min}$	Souřadnice vrcholů obr. 2.2		Zaokrouhlené $n_{\min}$	Zvětšení q_t při $n_{\min} - 1$
		q_t (y)	$n_{\min}$		
90	1,645	2,779	3,446	4	+76%
95	1,960	3,177	4,002	5	+56%
98	2,326	3,639	4,678	5	+43%
99	2,576	3,948	5,163	6	+38%
99,5	2,807	4,233	5,628	6	+34%
99,9	3,291	4,823	6,655	7	+28%

Rozsahy velmi malých výběrů odpovídající vrcholům závislosti $n_v = f(n, P)$ pro P 90 až 99,9% lze odhadnout pomocí rovnice (2.7); výsledky výpočtu se zaokrouhlují na nejbližší vyšší celé číslo

$$n_{\min} \approx 0,151u_p^2 + 1,2u_p + 1,07 \quad (2.7) \quad (\text{Použití } u_p \text{ viz pozn. na konci čl. 2.2.2.}).$$



Obr. 2.2: $q_t = f(n, P)$

Použití malých rozsahů výběrů je podmíněno absencí hrubých chyb a omylů. Výběr, zejména velmi malého rozsahu, musí být vždy prověřen a případné odlehlé hodnoty sledovaného znaku vyloučeny a nahrazeny novými daty nezátíženými hrubými chybami a omyly tak, aby byl doplněn počet jednotek nejméně na $n_{\min}$, resp. na $n \geq 5$. Způsob vylučování odlehlých výsledků je uveden v čl. 2.3.

Je nevýhodné volit příliš vysoké hodnoty P %. Například při $P = 99\%$ a rozsahu výběru $n = 10$ činí hodnota $q_t \approx 2,1$. Při volbě $P = 99,9\%$ a rozsahu $n = 10$ však činí hodnota $q_t \approx 3,0$. To znamená, že při zvětšení pravděpodobnosti pokrytí hodnoty μ o 0,9% se zvětší hodnota q_t o 45%! Proto je

nejčastější volbou $P = 95$ až 99% . Za malé výběry se považují výběry s rozsahy $n < 30$. **Pozn.:** Pro usnadnění výpočtů byly v rovnici (2.7) místo hodnot t_α použity kvantily u_p nezávislé na n ; tato úprava se promítla do konstant rovnice (2.7).

2.2.3 Velký rozsah výběru

Velké výběry jsou určeny pro přesnější odhady parametrů základního souboru. Při vyhledávání přijatelného rozsahu velkého výběru bylo využito dvoustranného konfidenčního intervalu (2.8), pokrývajícího hodnotu μ s $P = (1 - \alpha) \cdot 100\%$ pravděpodobnosti s kvantily normálního rozdělení u_p – viz tab. II, str. 42, $p = 1 - \alpha$

$$\bar{x} - \left(\frac{u_p}{\sqrt{n}} \right) \cdot \sigma \leq \mu \leq \bar{x} + \left(\frac{u_p}{\sqrt{n}} \right) \cdot \sigma \quad (2.8)$$

ze kterého lze šíři intervalu spolehlivosti $I_{\mu u}$ odvodit.

Parametr σ je pro každý základní soubor specifickou konstantou. Proto sledujme funkci 2.10.

$$I_{\mu u} = \left(\frac{2 \cdot u_p}{\sqrt{n}} \right) \cdot \sigma = q_u \cdot \sigma \quad (2.9)$$

$$\text{a tedy } q_u = \left(\frac{2 \cdot u_p}{\sqrt{n}} \right) \quad (2.10)$$

kde: σ = směrodatná odchylka základního souboru, kde odhad σ viz čl. 2.7.7

u_p = kvantil normálního rozdělení pro dvoustranný kritický obor. tab.II. s. 42

Na obr. 2.3 spojnice vrcholů n_v vyznačuje opět hodnoty $n_{\min}$, které nemají být podkročeny. Na rozdíl od malých výběrů velké výběry s použitím kvantilů normálního rozdělení vykazují menší citlivost hodnot $I_{\mu u}$ na podkročení hodnot $n_{\min}$. Vychází se z předpokladu přibližně normálního rozdělení chyb rozsahu $n \geq 30$. Na obr. 2.3 jsou grafy funkce $q_u = f(n, P)$. Průběh křivek je podobný jako pro malé výběry (posunutá rovnoběžná hyperbola).

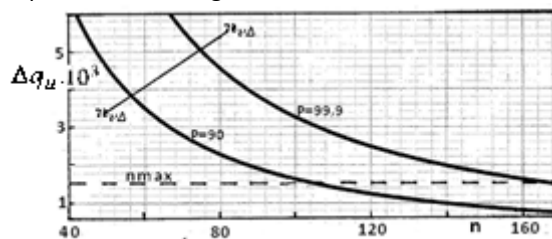
Při zvětšování velkého rozsahu výběru za účelem zvýšení přesnosti odhadů parametrů je účelné hodnotit efektivnost zvětšování rozsahu n . K tomu lze použít vztah (2.11)

$$\Delta q_{ui} = \left[\frac{2 \cdot u_p}{\sqrt{n_i}} - \frac{2 \cdot u_p}{\sqrt{n_i + 1}} \right] \quad (2.11)$$

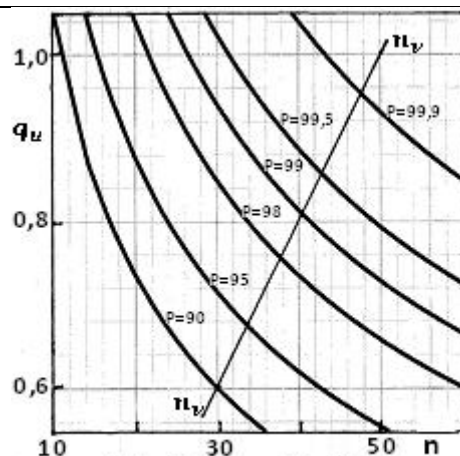
Tento vztah poskytuje difference Δq_{ui} odpovídající rozdílu hodnot q_u pro rozsahy n_i a $n_i + 1$.

Závislost těchto diferencí na rozsahu výběru n je opět různá pro různé pravděpodobnosti $P\%$ a je graficky znázorněna na obr. 2.4. Z průběhu závislosti pro $P = 90$ a $99,9\%$ je zřejmé, že pro n větší než odpovídá vrcholům křivek (označené jako $n_{v\Delta}$) se účinnost zmenšování Δq_{ui} rychle snižuje. To samozřejmě platí i pro ostatní závislosti nacházející se mezi těmito dvěma krajními.

S ohledem na snižující se efektivnost zvětšování rozsahu n by volby rozsahů výběrů neměly příliš překračovat hranici rozsahu $n_{v\Delta}$, není-li při výpočtu podmínkou dodržení maximální absolutní chyby odhadu předem dané velikosti a tím i výpočtu rozsahu n , např. podle vzorce (2.4). Za přijatelnou hranici maximálního rozsahu lze zvolit např. rozsah n , kdy difference činí pouze $\Delta q_{ui} = 1,5 \cdot 10^{-3}$. Malé změny hodnot Δq_u s růstem n nad $n_{\max}$ (čárkovaně) jsou zřejmé i z obr. 2.4, kdy je patrné zpomalení konvergence křivek k minimální hodnotě.



Obr. 2.4: Účinnost zvětšování rozsahu výběru



Obr. 2.3: $q_u = f(n, P)$

Tabulka 2.2 Přijatelné velké rozsahy výběrů normálního rozdělení

P%	90	95	98	99	99,5	99,9
u_p	1,645	1,960	2,326	2,576	2,807	3,291
$n_{\min}$	31	34	38	41	43	48
$n_{v\Delta}$	57	61	65	68	71	76
$n_{\max}$	106	119	134	143	152	169

Pro zvolené pravděpodobnosti P od 90 do 99,9% lze přijatelné velké rozsahy n odhadnout pomocí rovnic

$$n_{\min} \approx -0,738 \cdot u_p^2 + 14,35 \cdot u_p + 8,43 \quad (2.12)$$

$$n_{v\Delta} \approx -0,729 \cdot u_p^2 + 14,6 \cdot u_p + 34,93 \quad (2.13)$$

$$\text{Pro } \Delta q_u = 1,5 \cdot 10^{-3} \quad n_{\max} \approx -2,608 \cdot u_p^2 + 50,79 \cdot u_p + 29,41 \quad (2.14)$$

Vypočtené hodnoty n se opět zaokrouhlují na nejbližší vyšší celé číslo.

Rozsahy $n_{\min}$, $n_{v\Delta}$ z rovnic (2.12, 2.13) postačí jako přijatelné rozsahy i k odhadům intervalů spolehlivosti parametru σ podle vztahu (2.43) v čl. 2.7.11 s pomocí kvantilů rozdělení chí-kvadrát. Uvedené přijatelné rozsahy ani způsob jejich odvození však nelze použít při konstrukci tolerančních mezí. To je patrné z Jílkovy monografie [16].

Při plánování odběru vzorků za účelem hodnocení jejich analýz by pak rozsah výběru $n > n_{\max}$ měl být přesvědčivě zdůvodněn. Zpracování větších výběrů ($n > n_{\max}$) je přijatelné tehdy, jsou-li u vzorků získávány hodnoty znaků jakosti v důsledku zavedení pravidelné kontroly, nebo při geologickém průzkumu uhelné sloje, případně při registraci výsledků měření kontinuálním analyzátozem, takže náklady na hodnocení jsou dány prakticky pouze výdaji na počítačové zpracování dat.

Celková šíře intervalu spolehlivosti I_μ charakterizovaná součiny (2.5), resp. (2.9) potvrzuje použitelnost hodnot součinitelů $q(q_t, q_u)$ závislých na rozsahu výběru n a P% k odhadům přijatelných rozsahů výběrů. Volba rozsahů n v čl. 2.2.2 a 2.2.3 ukazuje, že existují určitá praktická omezení rozsahu výběru n daná jednak minimálním rozsahem $n_{\min}$, jednak přijatelným rozsahem $n_{v\Delta}$, a konečně volitelným rozsahem $n_{\max}$, jehož další zvětšování je neefektivní. To se však týká zmenšování vlivu chyb náhodných nikoliv soustavných; vlivy chyb soustavných na přesnost a správnost odhadů se zvětšením n nezměňují a vyžadují jiná opatření viz čl 2.5 a 2.9.1.6.1.

2.3 Vylučování odlehlých výsledků jednorozměrných dat

Vyskytne-li se v množině hodnot, zejména u výběru malého rozsahu, nápadně odlehlá hodnota znaku, je nutné její prověření, zda není zatížena hrubou chybou. Oba následující testy 2.3.1 a 2.3.2 mohou selhat v případě, že jsou přibližně stejně odlehlé obě okrajové hodnoty výběru, tj. x_1 i x_n . V takovém případě lze pokusně testovat výběr zmenšený o jednu okrajovou hodnotu a na základě výsledku vyloučit či ponechat okrajové hodnoty podle míry odlehlosti (viz čl. 2.3.3). Nelze však zmenšovat malý výběr pod rozsah $n_{\min}$ ($n < 5$). Jinak je nutno rozsah opět zvětšit na $n_{\min}$, resp. $n \geq 5$. Spolehlivější výsledky poskytují výběry s $n \geq 9$. Mimo následujících testů uvádí statistická literatura celou řadu dalších testů [17,18].

2.3.1 Dean - Dixonův test

Nejjednodušším neparametrickým testem pro odlehlé hodnoty malých výběrů je Dean - Dixonův test [7], vyžadující následující postup:

- Hodnoty souboru se seřadí podle velikosti $x_1 \leq x_2 \leq \dots \leq x_n$.
- Jako první se testuje hodnota nejvzdálenější od průměru pomocí kritérií

$$D_n = \frac{x_n - x_{n-1}}{x_n - x_1} \quad \text{respektive} \quad D_1 = \frac{x_2 - x_1}{x_n - x_1} \quad (2.15)$$

přičemž D_n je pro testování nejvyšších a D_1 nejnižších hodnot výběru.

- Vypočtená hodnota D_n nebo D_1 se porovná s tabelovanou kritickou hodnotou D_α pro zvolenou hladinu významnosti α a rozsah n z tabulky III. str. 43
- Vyazuje-li testovací kritérium (2.15) hodnotu větší nebo rovnou tabelované, lze odlehlou hodnotu x_n resp. x_1 ze souboru vyloučit, viz též čl. 2.3.3.
- Byl-li rozsah souboru před vylučováním minimální ($n = n_{\min}$, $n = 5$), je nutné po vyloučení odlehlé hodnoty rozsah výběru doplnit nejméně na $n_{\min}$, resp. $n \geq 5$.
- Testování končí, jsou-li hodnoty D_n , D_1 menší než příslušné kritické hodnoty D_α .

2.3.2 Grubbsův test

Grubbsův parametrický test [8] na odlehlost hodnot vyžaduje výpočet aritmetického průměru a kvadratické chyby hodnoceného výběrového souboru. Testovací kritéria mají podobu

$$G_n = \frac{x_n - \bar{x}}{S} \quad \text{resp.} \quad G_1 = \frac{\bar{x} - x_1}{S} \quad (2.16)$$

$$\text{kde: } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad S = \sqrt{\left[\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right] : n} \quad (2.17)$$

Při testování se postupuje analogicky jako v čl. 2.3.1, tj. po seřazení prvků podle velikosti se vypočítá hodnota G_n resp. G_1 a porovná s tabelovanou hodnotou G_α (tab. IV. str. 44). Je-li testovací kritérium G_n resp. G_1 vyšší nebo rovné tabelované kritické hodnotě G_α , lze odlehlou hodnotu x_n resp. x_1 vyloučit a zbylý výběr případně doplnit novými prvky na $n_{\min}$. Test končí, jsou-li hodnoty G_1 , G_n menší než příslušné hodnoty G_α pro zvolenou hladinu významnosti α .

2.3.3 Posouzení odlehlosti

Při posuzování odlehlosti hodnot výsledků, například analytických metod [9], se postupuje takto: V tabulkách se zjistí hladina α_0 , na níž je nalezena významnost shody či rozdílu testovacího kritéria od kritické hodnoty a rozhoduje se:

$\alpha_0 < 0,01$	→ hodnota je statisticky velmi významně odlehlá
$0,01 \leq \alpha_0 < 0,05$	→ hodnota je statisticky významně odlehlá
$0,05 \leq \alpha_0 < 0,10$	→ hodnota je statisticky málo odlehlá
$0,10 \leq \alpha_0$	→ hodnota není odlehlá

2.4 Test významnosti rozdílů $\bar{x} - \mu_0$

Studentův t-test významnosti rozdílů výběrového průměru $\bar{x}$ od předpokládané střední hodnoty μ_0 zahrnuje jako testovací kritérium

$$t = \frac{(\bar{x} - \mu_0)\sqrt{n}}{s} \quad (2.18)$$

kde: $\bar{x}$ = aritmetický průměr = $\frac{1}{n} \sum_{i=1}^n x_i$, pro $x_i = x_1, x_2, \dots, x_n$

μ_0 = očekávaná střední hodnota

s = výběrová směrodatná odchylka viz (2.38a,b).

Veličina t má Studentovo rozdělení t s $(v) = n - 1$ stupni volnosti.

1) Oboustranný kritický obor

Je-li $|t| > t_{\alpha}(v)$, je hodnota $\bar{x}$ významně odlišná od μ_0 na zvolené hladině významnosti α . Naopak, při $|t| < t_{\alpha}(v)$ je hodnota $\bar{x}$ nevýznamně odlišná od μ_0 na zvolené hladině významnosti α . Kritické hodnoty $t_{\alpha}(v)$ pro různé hladiny významnosti jsou v tabulce I.

2) Jednostranný kritický obor

a) Je-li $t > t'_{\alpha}(v)$, je hodnota $\bar{x}$ významně vyšší než hodnota μ_0 na zvolené hladině významnosti α (pravostranná alternativa $\mu > \mu_0$).

b) Je-li $t < -t'_{\alpha}(v)$, je hodnota $\bar{x}$ významně nižší než hodnota μ_0 na zvolené hladině významnosti α (levostranná alternativa $\mu < \mu_0$).

Hodnoty $t'_{\alpha}(v)$ v jednostranném kritickém oboru splňují relaci:

$$t'_{\alpha/2}(v) = t_{\alpha}(v)$$

Příklad: Průměrná zdánlivá hustota dřevěných briket má vykazovat hodnotu $\bar{d}_a^r \geq 1,00 \text{ g.cm}^{-3}$. Byl proměřen náhodný výběr vzorků briket o rozsahu $n = 11$ a u každé brikety byla stanovena zdánlivá hustota d_a^r v původním stavu

Č. vz.	1	2	3	4	5	6	7	8	9	10	11
d_a^r	0,97	0,93	1,01	1,00	0,93	0,99	0,97	0,92	0,99	1,01	0,92

Hodnoty potřebné k výpočtu testovacího kritéria podle vzorce statistiky (2.18):

$$\bar{x} = 0,967$$

$$\mu_0 = 1,00$$

$$n = 11$$

$$t = \frac{(0,967 - 1,00) \cdot \sqrt{11}}{0,0361} = -3,03$$

$$s = 0,0361$$

$$t'_{0,05}(10) = 1,812 \text{ viz tab. I}$$

Průměrná zdánlivá hustota náhodného výběru dřevěných briket je signifikantně nižší na zvolené hladině α , jak je zřejmé z nerovnosti: $t < -t'_{\alpha}(v) = -3,03 < -1,812$.

2.5 Vyhledávání soustavných chyb (odchylek)

K odhadům přítomnosti soustavných chyb, resp. odchylek lze použít párového testu. Párový test je pro dva porovnávané výběrové soubory, které jsou nějakým způsobem vzájemně vázány. To jsou např. provedená měření na každé jednotce výběru vzorků dvakrát, přičemž $x_i(x_1, x_2, \dots, x_n)$

jsou výsledky měření např. jednou referenční metodou (laboratoří, přístrojem, apod.) a $y_i (y_1, y_2, \dots, y_n)$ jsou výsledky ověřovaného měření téhož znaku druhou testovanou metodou (laboratoří, přístrojem, apod.) U náhodného výběru dvojic hodnot (x_i, y_i) stejného rozsahu, párově vázaných, se předpokládá, že pocházejí ze základních souborů normálního rozdělení. Pokud je na zvolené hladině významnosti α nalezena signifikantní přítomnost soustavných odchylek, je možné YOUNGHOVOU metodou (čl. 2.9.1.6.1) přibližně odhadnout, zda se ve výsledku kombinací různých chyb jedná o proporcionální či konstantní výslednou chybu a na základě její bližší charakteristiky se pokusit o identifikaci zdrojů jejího vzniku a případnou jejich eliminaci. Bližší specifikace soustavných chyb je náročná a ne vždy zcela úspěšná. Ve statistické literatuře je uvedena řada testů na indikaci soustavných chyb; dále se uvádějí pouze dva z nejpoužívanějších.

2.5.1 Párový t-test

Pro řadu dvojic porovnávaných hodnot (x_i -referenčních, y_i -ověřovaných) se vypočítá aritmetický průměr diferencí d_i a jejich druhých mocnin d_i^2 a testovací kritérium je pak

$$t = \bar{d} \cdot \sqrt{\frac{n-1}{\overline{d^2} - \bar{d}^2}} \quad (2.19)$$

$$\begin{aligned} \text{kde: } \bar{d} &= \frac{1}{n} \sum_{i=1}^n (y_i - x_i) \\ \overline{d^2} &= \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2 \end{aligned}$$

Při rozhodování v oboustranném kritickém oboru se testuje, zda $|t| < t_{\alpha}(v)$, kdy se průměrné hodnoty signifikantně neliší ($\bar{x} = \bar{y}$), či zda je $|t| > t_{\alpha}(v)$, kdy se průměrné hodnoty signifikantně liší ($\bar{x} \neq \bar{y}$) při zvolené hladině významnosti α . Hodnoty $t_{\alpha}(v)$ viz tab. I. s. 41.

Při rozhodování v jednostranném kritickém oboru se postupuje analogicky jako v případě 2) v článku 2.4, 2).

2.5.2 Znaménkový test

Stejná pravděpodobnost výskytu kladných i záporných odchylek v základním souboru distribuční funkce odpovídá symetrickému binomickému rozdělení. Střední hodnotou rozdělení je medián $\tilde{x}$ (čl. 2.6.1). Vztah mediánu k hodnotám náhodného výběru ($x_1, \dots, x_n$) lze vyjádřit vztahem $P(x < \tilde{x}) = P(x > \tilde{x}) = \frac{1}{2}$. Lze testovat difference hodnot výběru od mediánu $\tilde{x} - x_1$ až $\tilde{x} - x_n$. Znaménkový test se nejčastěji používá jako párový test.

U výběrů se nejdříve zjistí rozdíly $\pm d_i = (y_i - x_i)$ včetně znamének a jejich počtů $\sum d_i^+$, $\sum d_i^-$. Vzniknou-li nějaké nulové rozdíly d_i^0 , provede se jejich úprava tak, že polovině z nich se přisoudí kladná znaménka a druhé polovině záporná. Při lichém počtu d_i^0 se přidá jedna difference potřebného znaménka. Za rozsah n se vezme počet všech znamének $n = \sum n_i^+ + \sum n_i^-$. V tabulce V. str. 45 - 46 jsou kritické hodnoty k_1 a k_2 pro různé rozsahy n a různé hladiny významnosti α . Pokud se zjištěné počty znamének n^+ a n^- při zvolené hladině významnosti α a odpovídajícím rozsahu n nalézají uvnitř mezi kritickými hodnotami k_1 a k_2 , je výběr zatížen pouze náhodnými chybami – odchylkami. Je-li menší z hodnot počtu znamének n^+ , n^- rovna nebo menší než k_1 a větší z hodnot počtu znamének n^+ , n^- rovna nebo větší než k_2 , lze usuzovat na

přítomnost systematických chyb – odchylek v hodnoceném výběru při zvolené hladině významnosti α . Obvykle se volí $\alpha = 0,05$ až $0,01$. Neparametrický znaménkový test, podobně jako všechny neparametrické testy, je při malém rozsahu souboru málo citlivý. Zlepšené citlivosti lze dosáhnout zvětšením rozsahu výběrů na hodnoty zhruba odvozené ze závislosti $k_2 : k_1 = f(n)$, tak, aby přijatelné hodnoty rozsahů byly za vrcholy hyperbolických křivek těchto funkcí, kdy poměry $k_2 : k_1$ jsou < 5 . Tyto rozsahy činí pro $\alpha = 0,2$ $n \geq 9$; pro $\alpha = 0,1$ $n \geq 13$; $\alpha = 0,05$ $n \geq 15$; $\alpha = 0,02$ $n \geq 19$; $\alpha = 0,01$ $n \geq 21$. Znaménkový test se uplatní i v případě, kdy je soustavná chyba tak malá, že by se při použití párového t-testu nemusela projevit jako signifikantní. Je zřejmé, že znaménkový test je vhodný pro větší výběry.

2.6 Střední hodnoty, charakteristiky úrovně (polohy)

Existuje mnoho druhů středních hodnot. Ve statistice jsou to charakteristiky, které zevšeobecnují velikost hodnot určitého statistického znaku u všech prvků daného souboru jedním číslem, které zastupuje jednotlivé hodnoty uvažované řady. Průměry jsou odvozeny z hodnot všech prvků řady. Rozsahy souborů pro odhady středních hodnot lze určit podle čl. 2.2.2 a 2.2.3. Střední hodnoty lze určit jako nevážené či jako vážené, přičemž váhami mohou být četnosti, hmotnosti, procenta apod. Medián a modus jsou odvozeny jen z některých charakteristických hodnot řady.

2.6.1 Medián $\tilde{x}$

Medián je prostředním prvkem – jednotkou řady hodnot jednotek výběru, seřazených podle velikosti. Hodnota jednotky vyznačené jako medián je závislá na pořadí jednotek z hlediska hodnoty zkoumaného znaku. Medián je tedy citlivý k pořadí, nikoliv k hodnotám znaku. Pro $n > 2$ není závislý na krajních hodnotách. Je-li v řadě lichý počet jednotek, je medián hodnotou s pořadím $(n+1):2$. Je-li v řadě sudý počet jednotek, je medián hypotetickou hodnotou, a to aritmetickým průměrem dvou hodnot, které jsou mediánu nejbližší. Je to tudíž aritmetický průměr dvou hodnot znaku příslušejícím jednotkám s pořadím $[(n+1):2] - 0,5$ a $[(n+1):2] + 0,5$.

2.6.2 Aritmetický průměr $\bar{x}$

Prostý aritmetický průměr je výběrovým úhrnem hodnot znaku, děleným rozsahem výběru. Je nestranným, konzistentním a velice vydatným odhadem parametru μ základního souboru normálního rozdělení.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.21)$$

Součet odchylek jednotlivých hodnot x_i od aritmetického průměru je roven nule. Součet čtverců odchylek od aritmetického průměru je vždy menší než součet čtverců odchylek těchto hodnot od jiné, libovolné hodnoty. Aritmetický průměr součtů znaků x a y se rovná součtu aritmetických průměrů těchto znaků: $\overline{x \pm y} = \bar{x} \pm \bar{y}$.

2.6.3 Vážený aritmetický průměr $\overline{x_v}$

Při výpočtu $\overline{x_v}$ se jednotlivým hodnotám znaku x_i přisuzuje různá váha v_i .

$$\overline{x_v} = \frac{\sum_{i=1}^k x_i v_i}{\sum_{i=1}^k v_i} \quad (2.22)$$

kde: k = počet variant znaku při $i = 1, 2, \dots, k$.

2.6.4 Harmonický průměr $\overline{x}_h$

Nevážený harmonický průměr je převratnou hodnotou aritmetického průměru převratných hodnot znaku

$$\overline{x}_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad (2.23)$$

Harmonický průměr se používá tehdy, je-li lineárně aditivní reciproká hodnota znaku, např. reciproká hodnota hustoty (zdánlivé, skutečné, atd.).

2.6.5 Vážený harmonický průměr $\overline{x}_{hv}$

Vážený harmonický průměr se vypočítá podle vztahu

$$\overline{x}_{hv} = \frac{\sum_{i=1}^k v_i}{\sum_{i=1}^k \frac{v_i}{x_i}} \quad (2.24)$$

Jednotlivým různým hodnotám znaků x_i se přisuzuje opět různá váha.

2.6.6 Geometrický průměr $\overline{x}_g$

Geometrický průměr je dán vztahem

$$\overline{x}_g = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} \quad (2.25)$$

jehož logaritmováním se dostane

$$\lg \overline{x}_g = \frac{\sum_{i=1}^n \lg x_i}{n} \quad (2.26)$$

Geometrický průměr se tedy určuje tak, že se vypočítá aritmetický průměr logaritmů hodnot x_i a antilogaritmováním se získá hledaný geometrický průměr. K použití geometrických průměrů dochází např. v analýze časových řad, při popisu levostranně nesouměrných rozložení četností pomocí transformace nesouměrného rozložení logaritmováním v souměrnější rozložení, které se snáze zkoumá, viz např. lognormální rozdělení [13], v korelačním počtu atd.

2.6.7 Vážený geometrický průměr $\overline{x}_{gv}$

Vážený geometrický průměr je dán vztahem

$$\overline{x}_{gv} = \sqrt[v]{\prod_{i=1}^k x_i^{v_i}} = \sqrt[v]{x_1^{v_1} \cdot x_2^{v_2} \cdot \dots \cdot x_k^{v_k}} \quad \text{přičemž } v = \sum_{i=1}^k v_i \quad (2.27)$$

a logaritmováním se dostane

$$\lg \overline{x_{gv}} = \frac{\sum_{i=1}^k v_i \lg x_i}{v} \quad (2.28)$$

2.6.8 Kvadratický průměr $\overline{x_k}$

Uvažuje-li se kvantitativní statistický znak x , který nabývá n hodnot $x_1, x_2, \dots, x_n$, potom je kvadratický průměr definován jako

$$\overline{x_k} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}} \quad (2.29)$$

Výsledku se přisuzuje kladné znaménko. Kvadratický průměr se užívá zejména při měření kumulace chyb - variability při základních operacích, např. sčítání a odčítání x, y, z , viz čl. 2.7.12.

2.6.9 Vážený kvadratický průměr $\overline{x_{kv}}$

Vážený kvadratický průměr je dán vztahem

$$\overline{x_{kv}} = \sqrt{\frac{\sum_{i=1}^k x_i^2 v_i}{\sum_{i=1}^k v_i}} \quad (2.30)$$

2.6.10 Modus $\hat{x}$

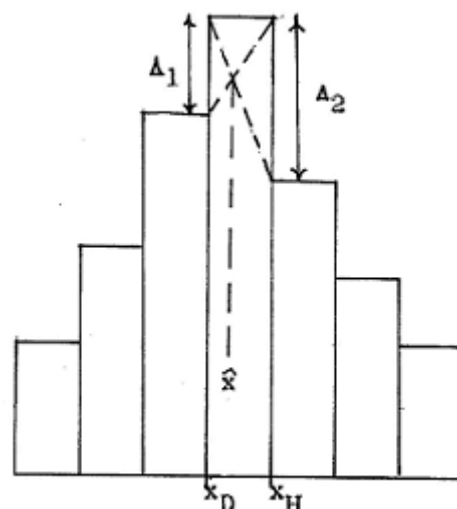
U velkých výběrů se zkoumají též intervalová rozložení četností s uvedením rozpětí tříd, průměrných hodnot a četností tříd, buď tabelárně, nebo formou sloupcového histogramu nebo polygonu.

Při určování počtu tříd v histogramu lze použít Sturgesova vzorce, kde počet tříd $\leq 1 + 3,3 \cdot \lg n$. Modus je pak nejčastěji se vyskytující střední hodnota ve zkoumaném souboru.

Určení modu nečiní potíže u jednovrcholového rozložení četností a zejména u veličin s opakujícími se hodnotami. Podmínkou je, aby v blízkosti nejčetnějšího intervalu byly stejné intervaly. Modus se pak vypočítá podle vztahu

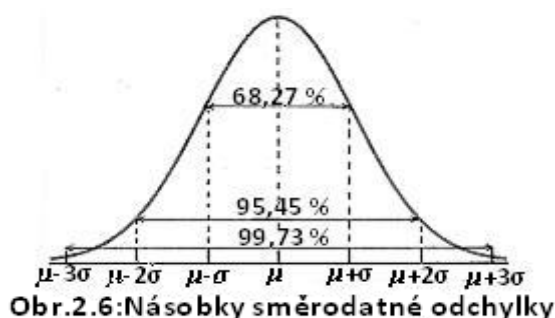
$$\hat{x} = \frac{x_D \cdot \Delta_2 + x_H \cdot \Delta_1}{\Delta_1 + \Delta_2} \quad (2.31)$$

kde: x_D = dolní mez nejčetnějšího intervalu,
 x_H = horní mez nejčetnějšího intervalu,
 Δ_1 = rozdíl mezi největší četností a četností intervalu předcházejícího nejčetnější interval.
 Δ_2 = rozdíl mezi největší četností a četností příslušné intervalu následujícímu po nejčetnějším intervalu.



Obr.2.5: Schema výpočtu modu $\hat{x}$

Struktura výpočtu polohy modu intervalového rozložení četností spojité veličiny je znázorněna na obr. 2.5. Při dostatečně velkém rozsahu n poskytuje histogram přibližnou informaci o symetrii či asymetrii rozložení četností. Přesnější informace o symetrii viz čl. 2.8.1.



Obr. 2.6: Násobky směrodatné odchylky



Obr. 2.7: Hladina významnosti α

2.7 Charakteristiky variability

Jak se jednotlivé hodnoty znaku v souboru vzájemně liší a jak se liší od středu, vyjadřují parametry variability. Hodnocení variability je důležité pro posouzení přesnosti a spolehlivosti charakteristik úrovně a pro získání zhuštěné informace o měnlivosti hodnot souboru. Střední hodnoty teprve ve spojení s ukazateli variability získávají potřebnou vypovídací hodnotu.

2.7.1 Variační rozpětí R

Nejjednodušší mírou měnlivosti hodnot znaku ve výběru je variační rozpětí

$$R = x_{\max} - x_{\min} \quad (2.32)$$

Velká citlivost R k extrémním okrajovým hodnotám a necitlivost k ostatním hodnotám je značnou nevýhodou R . Variační rozpětí se přibližně rovná šesti směrodatným odchylkám „pravidlo šesti sigma“ $R \approx 6\sigma$. Podstatu tohoto pravidla a význam parametru σ objasňuje obr. 2.6, kde je zřejmé, že rozpětí $\mu \pm 3\sigma$ základního souboru vymezuje pod Gausovou křivkou 99,73 % plochy, což se prakticky blíží 100 %. Rozpětí $\mu \pm \sigma$ (68,27 %) je pak dáno vzdáleností dvou inflexních bodů v ramenech Gaussovy křivky.

Podrobnější přehled vztahu mezi hodnotou intervalu $1 - \alpha = p$ a kvantily normálního rozdělení u_p , které prezentují pravděpodobnost výskytu chyby v mezích $\pm u_p$ násobku směrodatné odchylky σ , je v tabulce II. s. 41. Šíře intervalu spolehlivosti $1 - \alpha$ v oboustranné alternativě odhadu konfidenčních mezí pro parametr μ je na obr. 2.7 vymezena šrafovanými plochami odpovídajícími hladině významnosti α , resp. $2 \cdot (\frac{1}{2}\alpha)$ od $-\infty$ do u_1 a od u_2 do $+\infty$. I když tabulky IIa,b, s. 41 umožňují výběr hodnot $1 - \alpha$ a odpovídajících kvantilů u_p pro nejrůznější úkoly v širokém rozpětí, obvykle používáme nejčastěji $\alpha = 0,05$ až $0,01$, resp. $1 - \alpha = 0,95$ až $0,99$.

2.7.2 Průměrná odchylka $|\bar{d}|$

Průměrná odchylka je aritmetickým průměrem součtu absolutních hodnot jednotlivých odchylek znaku od zvolené libovolné hodnoty, nejčastěji aritmetického průměru

$$|\bar{d}| = \frac{1}{n} \sum_{i=1}^n |d_i| \quad (2.33)$$

Průměrná odchylka se používá výhradně pro přímé vyjádření úrovně variability.

2.7.3 Odhad rozptylu základního souboru $\hat{\sigma}^2$

Nestranný odhad rozptylu základního souboru normálního rozdělení se provede výpočtem s použitím počtu stupňů volnosti $\nu = n - 1$ podle vztahu

$$\hat{\sigma}^2 \equiv s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2 \right] \quad (2.35)$$

kde: s^2 = výběrový rozptyl, který označujeme malým písmenem s^2 , na rozdíl od rozptylu statistického souboru, který označujeme velkým písmenem S^2 , viz čl. 2.7.13.

2.7.4 Odhad disperze $\hat{D}(s^2)$

Rozptyl (s^2), disperze (D), variance (Var), jsou synonyma pro odlišení rozptylů různých parametrů. Tudíž nestranný odhad disperze parametru s^2 je dán vztahem

$$\hat{D}(s^2) = \frac{2s^4}{(n-1)+2} \quad [13] \quad (2.36)$$

2.7.5 Dvoustranný konfidenční interval rozptylu σ^2

Meze dvoustranného konfidenčního intervalu rozptylu σ^2 základního souboru, normálního rozdělení, se určí podle vztahu

$$\left(\frac{n-1}{\chi_{(1+p)/2}^2(\nu)} \right) \cdot s^2 \leq \sigma^2 \leq \left(\frac{n-1}{\chi_{(1-p)/2}^2(\nu)} \right) \cdot s^2 \quad (2.37)$$

kde $(\nu) = (n - 1)$.

Kvantily rozdělení χ^2 (chí kvadrát) viz tabulka VI. s. 47 - 48. Výběr kvantilů χ^2 viz komentář k tabulce VI.

K odhadu se použijí velké výběry s rozsahy uvedenými v tab. 2.2.

2.7.6 Výběrová směrodatná odchylka s

Vychýlený odhad výběrové (střední, kvadratické, směrodatné) odchylky jednoho měření je kladnou druhou odmocninou výběrového rozptylu, tudíž

$$s = +\sqrt{s^2} \quad (2.38a)$$

kde: s^2 viz (2.35).

Směrodatná odchylka pro případ paralelních dvojic stanovení znaku na w vzorkách se odhadne pomocí vztahu

$$s = \sqrt{\frac{1}{2w} \sum_{j=1}^w d_j^2} \quad \text{pro } j = 1, 2, \dots, w \quad (2.38b)$$

kde: w = počet dvojic paralelních stanovení (počet vzorků)

d_j = rozdíl mezi j -tou paralelní dvojicí stanovení $d_j = x_{j1} - x_{j2}$

2.7.7 Odhad směrodatné odchylky základního souboru $\hat{\sigma}$

Nestranný odhad směrodatné odchylky základního souboru normálního rozdělení je dán součinem

$$\hat{\sigma} = s \cdot k_v \quad (2.39)$$

$$\text{kde } k_v \cong \frac{(2:v)+1}{7,5+4v} + 1$$

Výpočet k_v pro $v = (n - 1) \geq 10$ platí na 4 desetinná místa.

Tabulka 2.3 Hodnoty k_v pro $v = 1$ až 10

v	1	2	3	4	5	6	7	8	9	10
k_v	1,2533	1,1284	1,0854	1,0638	1,0509	1,0424	1,0362	1,0317	1,0281	1,0253

Součinitelé k_v jsou významnější pro malé výběry.

2.7.8 Odhad disperze $\hat{D}(s \cdot k_v)$

Pro tento nestranný odhad disperze platí vztah

$$\hat{D}(s \cdot k_v) = (k_v^2 - 1) \cdot s^2 \quad [13] \quad (2.40)$$

2.7.9 Odhad směrodatné odchylky aritmetického průměru $s_{\bar{x}}$

Odhad směrodatné odchylky aritmetického průměru $\bar{x}$ je dán vztahem

$$s_{\bar{x}} = s : \sqrt{n} \quad (2.41)$$

2.7.10 Relativní směrodatná odchylka s_{rel}

Je jednou z nejlepších měr relativní variability. Uvádí se někdy vynásobena 100 v %, starším názvem též „variační koeficient“. Je dána vztahem

$$s_{rel} = s : \bar{x} \quad (2.42)$$

2.7.11 Interval spolehlivosti pro parametr σ

Dvoustranné konfidenční meze směrodatné odchylky základního souboru vyjadřuje nerovnost

$$\sqrt{\frac{n-1}{\chi_{(1+p):2}^2(v)}} \cdot s < \sigma < \sqrt{\frac{(n-1)}{\chi_{(1-p):2}^2(v)}} \cdot s \quad (2.43)$$

Hodnoty χ^2 viz tab. VI. str. 47 - 48. Přijatelné rozsahy výběrů pro výpočet intervalu – lze použít hodnoty rozsahů n z tab. 2.2.

2.7.12 Hromadění chyb při základních operacích

Hromadění chyb při základních početních operacích s hodnotami zatíženými chybou, jsou-li navíc dvě hodnoty, např. x , y , vzájemně závislé a míra jejich závislosti je charakterizována korelačním koeficientem r [10] (r viz čl. 2.9.1.1). Nejsou-li hodnoty x , y vzájemně závislé, odpadá třetí člen pravé strany rovnic (2.44) až (2.47).

FUNKCE HROMADĚNÍ CHYB [10]

$$z = x + y \quad s_z^2 = s_x^2 + s_y^2 + [2rs_x s_y] \quad (2.44)$$

$$z = x - y \quad s_z^2 = s_x^2 + s_y^2 - [2rs_x s_y] \quad (2.45)$$

$$z = x \cdot y \quad \left(\frac{s_z}{z}\right)^2 = \left(\frac{s_x}{x}\right)^2 + \left(\frac{s_y}{y}\right)^2 + \left[2r \left(\frac{s_x}{x}\right) \left(\frac{s_y}{y}\right)\right] \quad (2.46)$$

$$z = \frac{x}{y} \quad \left(\frac{s_z}{z}\right)^2 = \left(\frac{s_x}{x}\right)^2 + \left(\frac{s_y}{y}\right)^2 - \left[2r \left(\frac{s_x}{x}\right) \left(\frac{s_y}{y}\right)\right] \quad (2.47)$$

2.7.13 Rozptyl statistického souboru S^2

Na rozdíl od výběrového rozptylu s^2 používá vztah pro odhad statistického rozptylu S^2 pouze rozsah výběru n . K rozlišení obou rozptylů se u statistického rozptylu používá označení velké S^2 , na rozdíl od výběrového souboru s malým s^2 zahrnujícího ve vzorci stupně volnosti (v) = ($n - 1$), viz (2.35), tedy

$$S^2 = \left[\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right] : n \quad (2.48)$$

Rozptyl S^2 (disperse **D**, variance **Var**) má vlastnosti, kterých se využívá ke zpřehlednění vztahů a usnadnění výpočtů např. při hodnocení závislostí. Vzorec (2.48) je jedním z výpočtových tvarů rozptylu, které zjednodušují výpočet, neboť odpadá počítání odchylek jednotlivých hodnot od aritmetického průměru, jak to vyžaduje tvar $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$.

2.7.14 Vážený rozptyl statistického souboru S_v^2

U rozdělení četností znaku odhadujeme rozptyl formou vážených průměrů, kdy vahami jsou četnosti n_i při k variantách znaku, kde $i = 1, 2, \dots, k$, tudíž

$$S_v^2 = \frac{\sum_{i=1}^k x_i^2 \cdot n_i}{\sum_{i=1}^k n_i} - \left[\frac{\sum_{i=1}^k x_i \cdot n_i}{\sum_{i=1}^k n_i} \right]^2 = \frac{\sum_{i=1}^k x_i^2 \cdot n_i - \bar{x} \cdot \sum_{i=1}^k x_i \cdot n_i}{\sum_{i=1}^k n_i} \quad (2.49)$$

2.7.15 Celkový vážený rozptyl z rozptylů skupinových souborů

Rozptyl hodnot znaku v souboru složeném z více dílčích souborů se rovná součtu váženého aritmetického průměru rozptylů dílčích souborů a rozptylu dílčích aritmetických průměrů, tedy

$$S_v^2 = \frac{\sum_{i=1}^k S_i^2 \cdot n_i}{\sum_{i=1}^k n_i} + \left[\frac{\sum_{i=1}^k \bar{x}_i^2 \cdot n_i}{\sum_{i=1}^k n_i} - \bar{x}^2 \right] \quad (2.50)$$

kde: S_i^2 = rozptyl i -tého souboru,
 n_i = rozsah i -tého souboru, $i = 1, 2, \dots, k$

$\bar{x}_i$ = aritmetický průměr i-tého souboru,

$\bar{x}$ = celkový aritmetický průměr.

Vztah (2.50) umožňuje výpočet celkového váženého rozptylu i bez znalosti původních hodnot znaku. Postačí, jsou-li známy rozsahy jednotlivých skupin n_i , skupinové aritmetické průměry $\bar{x}_i$ a skupinové rozptyly S_i^2 .

2.8 Ověřování normality rozdělení četností

Dosud zmíněné charakteristiky úrovně a variability nepostačují, má-li být hodnocena normalita rozdělení četností. Mohou nastat případy, kdy parametry úrovně a variability u tří různých výběrů budou totožné, včetně rozsahu, avšak jejich rozdělení se budou zřetelně lišit v šikmosti a špičatosti.

2.8.1 Charakteristika šikmosti

Jednou z charakteristik šikmosti vystihující asymetrii rozložení četností je tzv. koeficient šikmosti (asymetrie) K_a (též výběrová šikmost). Vyhovuje všem hlavním požadavkům kladeným na míry šikmosti. Je dán vztahem

$$K_a = m_3 : (S_v^2 \cdot \sqrt{S_v^2}) \quad (2.51)$$

kde třetí centrální moment m_3 je dán vztahem

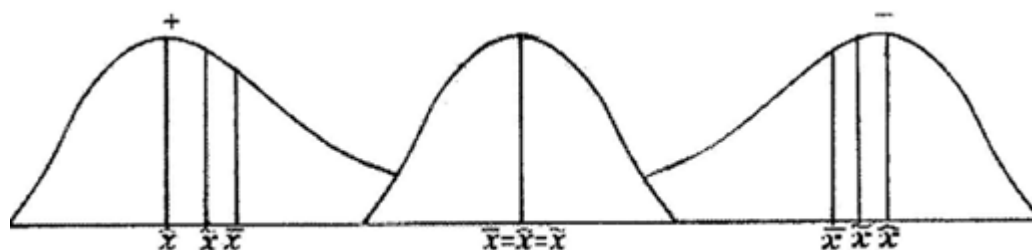
$$m_3 = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^3 \cdot n_i = \frac{\sum_{i=1}^k x_i^3 \cdot n_i}{\sum_{i=1}^k n_i} - 3\bar{x} \frac{\sum_{i=1}^k x_i^2 \cdot n_i}{\sum_{i=1}^k n_i} + 2\bar{x}^3 \quad (2.52)$$

U symetrických rozdělení je koeficient $K_a = 0$; je-li kladný, jedná se o rozdělení s nakupením hodnot vlevo od středu (zešikmení do leva), je-li záporný, je rozdělení s nakupením hodnot vpravo od středu (zešikmení doprava), viz obr.2.8. Absolutní hodnota K_a je tím větší, čím je dané rozdělení nesousměrnější.

Test založený na šikmosti [11] zamítne hypotézu o normalitě v případě, že

$$\frac{|K_a|}{\sqrt{\text{var } K_a}} \geq u_{\alpha/2} \quad (2.53)$$

kde: $\text{var } K_a = \frac{6 \cdot (n-2)}{(n+1) \cdot (n+3)}$



Obr.2.8: Vliv symetrie a šikmosti na polohu středních hodnot $\bar{x}, \tilde{x}, \ddot{x}$

2.8.2 Charakteristika špičatosti

Charakteristika vystihující špičatost, resp. plochost rozdělení četností se nazývá koeficient špičatosti (excesu) K_e (též výběrová špičatost). Vyhovuje všem hlavním požadavkům kladeným na míry špičatosti. Umožňuje měření koncentrace hodnot znaku kolem aritmetického průměru. Je dán vztahem

$$K_e = [m_4 : (S_v^2)^2] - 3 \quad (2.54)$$

kde čtvrtý centrální moment m_4 je dán vztahem

$$\begin{aligned} m_4 &= \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^4 \cdot n_i = \\ &= \frac{\sum_{i=1}^k x_i^4 \cdot n_i}{\sum_{i=1}^k n_i} - 4\bar{x} \frac{\sum_{i=1}^k x_i^3 \cdot n_i}{\sum_{i=1}^k n_i} + 6\bar{x}^2 \frac{\sum_{i=1}^k x_i^2 \cdot n_i}{\sum_{i=1}^k n_i} - 3\bar{x}^4 \end{aligned} \quad (2.55)$$

Koeficient K_e je pro normální rozdělení $K_e = 0$. Jeho kladné hodnoty znamenají, že hodnocené rozdělení je ve srovnání s normálním rozdělením špičatější; záporné hodnoty K_e svědčí o rozdělení plošším.

Test založený na špičatosti [11] zamítne hypotézu o normalitě, když

$$\frac{|K_e + (\frac{6}{n+1})|}{\sqrt{\text{var } K_e}} \geq u_{\alpha/2} \quad (2.56)$$

$$\text{kde: } \text{var } K_e = \frac{24n \cdot (n-2) \cdot (n-3)}{(n+1)^2 \cdot (n+3) \cdot (n+5)}$$

2.9 Lineární regresní závislosti

2.9.1 Regresní přímka

U přímky $\eta = \underline{a} + \underline{b} \cdot x$ jsou regresní odhady parametrů závislé na tom, která z proměnných dvourozměrného souboru x, y je bezchybná či zatížená chybami a podle jakých kritérií se hodnotí vliv reziduí e na odhad parametrů zvoleného regresního modelu. Experimentálními body se souřadnicemi x_i, y_i proložená vyrovnávací přímka je charakterizována empirickým vztahem

$$Y_i = a + b \cdot x_i + e_i \quad (2.57)$$

kde: Y_i = regresní odhad i -té závislé proměnné,

a = odhad aditivní konstanty posunutí (průsečíku regresní přímky s osou y),

b = odhad regresního koeficientu (směrnice regresní přímky),

x_i = i -tá hodnota nezávislé proměnné,

$e_i = (y_i - Y_i)$ = reziduum = soubor všech nezahrnutých veličin a chyb.

Následně se uvažují tři z variant regresních odhadů metodou nejmenších čtverců (MNČ) podle toho, u které z proměnných se předpokládá zatížení chybami vyjádřenými směrodatnými odchylkami proměnných základního souboru - rozuměj nepřesností měření experimentálních dat, za předpokladu bezchybné volby lineárního regresního modelu:

1) $\sigma_x = 0, \sigma_y \neq 0$, $\rightarrow$ nejčastěji používaná varianta, kdy nezávislé proměnné x jsou bezchybné hodnoty a závislé proměnné y jsou zatížené chybami. Používá se klasický model MNČ čl. 2.9.1.2, kdy minimalizace reziduí probíhá ve směru osy y .

2) $\sigma_x \neq 0, \sigma_y = 0$, $\rightarrow$ nejméně uvažovaný případ, uvedený pro úplnost a srovnání, kdy proměnné x jsou zatížené chybami a proměnné y jsou bezchybné. Výpočet MNČ viz čl. 2.9.1.3, kdy minimalizace reziduí je realizována ve směru osy x .

3) $\sigma_x \neq 0, \sigma_y \neq 0$, $\rightarrow$ ve skutečnosti často se vyskytující reálná varianta, kdy jsou obě proměnné zatíženy chybami. Minimalizace vzdáleností experimentálních bodů od proložené regresní přímky metodou MNČ se realizuje na kolmicích k regresní přímce. Výpočet viz čl. 2.9.1.4.

Tento model poskytuje z uvedených tří modelů nejnižší hodnoty reziduí a nejspolehlivější odhad rovnice regresní přímky.

Realizace těchto tří modelů u jednoho souboru dat poskytuje svazek tří přímek s průsečíkem v bodě $\bar{x}, \bar{y}$ (těžiště). Přímka podle třetí varianty se nachází mezi druhými dvěma, nikoliv obecně uprostřed [1]. Pro odhady parametrů regresních závislostí a, b s vyhovující přesností je důležité, aby výběr dat obsahoval v reálné oblasti co největší variační rozpětí proměnných x_i , za podmínky, že zejména okrajové hodnoty proměnných (pro $x_{\min}, x_{\max}$) nejsou zatíženy hrubými chybami. Potřebné minimální rozsahy výběrů pokryjí hodnoty $n_v(\min)$ uvedené v tabulce 2.2 čl. 2.2.3. Minimální rozsahy se rozumí po vyloučení vybočujících dat z výběru. Jsou-li náhodnostní a závislostní složky v daném výběru nekorelované, rozptyl chyb konstantní a mají-li normálně rozdělené náhodné chyby prakticky nulovou střední hodnotu, je za absence extrémních chyb nejlepším odhadem apriorní lineární regresní funkce empirická regresní funkce stanovená MNČ. Přesnější odhady parametrů získáme zpracováním neuspořádaných dat; méně přesné, avšak méně pracné (u rozsáhlých výběrů) je zpracování dat z tabelovaných podmíněných průměrů nebo z korelační tabulky, případně s použitím vážených dat. Pro všechny tři uvedené varianty odhadů parametrů regresních přímek pomocí experimentálních dat výběru se nejdříve vypočítají potřebné sumace a další hodnoty - pro zjednodušení jsou vynechány meze sumací, které se zde vždy rozumí $\sum_{i=1}^n$ kde $i = 1, 2, \dots, n$.

- | | |
|---------------|--|
| a) Sumace | $\sum x_i, \sum y_i, \sum x_i^2, \sum y_i^2, \sum x_i y_i$, rozsah výběru n |
| b) Průměry | $\bar{x} = \frac{1}{n} \sum x_i, \bar{y} = \frac{1}{n} \sum y_i, \overline{x^2} = \frac{1}{n} \sum x_i^2, \overline{y^2} = \frac{1}{n} \sum y_i^2, \overline{xy} = \frac{1}{n} \sum x_i \cdot y_i$, |
| c) Variance | $S_x^2 = \overline{x^2} - \bar{x}^2$
$S_y^2 = \overline{y^2} - \bar{y}^2$ |
| d) Kovariance | $S_{yx} = S_{xy} = \overline{xy} - \bar{x} \cdot \bar{y}$ |

2.9.1.1 Korelační koeficient r_{yx}

Předpokládá-li se mezi dvěma znaky výběru lineární korelace, pak mírou těsnosti této závislosti je empirický korelační koeficient daný vztahem

$$r_{yx} = r_{xy} = S_{yx} : \sqrt{S_x^2 \cdot S_y^2} \quad (2.58)$$

Hodnota r_{yx} se pohybuje v uzavřeném intervalu $\langle -1, 1 \rangle$. V absolutní hodnotě se blíží tím více jedné, čím je závislost těsnější. Kladných hodnot nabývá, jde-li o přímou závislost, záporných, jedná-li se o nepřímou závislost. Čtverec $r_{yx}^2 \cdot 100 = \rho$ nazývaný koeficient determinace ukazuje, z kolika % lze změny závisle proměnné vysvětlit zvolenou funkcí. Existenci závislosti se zvolenou pravděpodobností lze ověřit pomocí kritické hodnoty r_α (viz tab. VII. s. 49). Je-li $r_\alpha < |r_{yx}|$ pro $v = n - 2$, lze konstatovat, že existuje lineární korelační závislost mezi proměnnými x , y hodnoceného výběru dat se zvolenou pravděpodobností $P(1-\alpha) \cdot 100\%$.

2.9.1.2 Regresní přímka $Y = a_1 + b_1 \cdot x$ podle varianty $\sigma_x = 0, \sigma_y \neq 0$

Minimalizace čtverců odchylek probíhá ve směru osy y .

$$\text{Regresní koeficient} \quad b_1 = S_{yx} : S_x^2 \quad (2.59)$$

$$\text{Konstanta posunutí} \quad a_1 = \bar{y} - b_1 \cdot \bar{x} \quad (2.60)$$

$$\text{Reziduální směrodatná odchylka} \quad s_{e1} = \sqrt{S_y^2 \cdot (1 - r_{yx}^2) \frac{n}{n-2}} \quad (2.61)$$

$$\text{Směrodatná odchylka regresního koeficientu} \quad s_{b1} = s_{e1} \cdot \sqrt{1 : (S_x^2 \cdot n)} \quad (2.62)$$

$$\text{Směrodatná odchylka konstanty posunutí} \quad s_{a1} = s_{e1} \cdot \sqrt{\bar{x}^2 : (S_x^2 \cdot n)} \quad (2.63)$$

2.9.1.3 Regresní přímka $Y = a_2 + b_2 \cdot x$ podle varianty $\sigma_x \neq 0, \sigma_y = 0$

Varianta MNČ je uplatněna ve směru osy x .

$$\text{Regresní koeficient} \quad b_2 = S_y^2 : S_{yx} \quad (2.64)$$

$$\text{Konstanta posunutí} \quad a_2 = \bar{y} - b_2 \cdot \bar{x} \quad (2.65)$$

$$\text{Reziduální směrodatná odchylka} \quad s_{e2} = \sqrt{S_x^2 \cdot (1 - r_{yx}^2) \frac{n}{n-2}} \quad (2.66)$$

Směrodatná odchylka regresního koeficientu

$$s_{b2} = s_{e2} \cdot \sqrt{1 : (S_y^2 \cdot n)} \quad (2.67)$$

Směrodatná odchylka konstanty posunutí

$$s_{a2} = s_{e2} \cdot \sqrt{x^2 : (S_x^2 \cdot n)} \quad (2.68)$$

2.9.1.4 Regresní přímka $Y = a_3 + b_3 \cdot x$ podle varianty $\sigma_x \neq 0, \sigma_y \neq 0$

MNČ je uplatněna na kolmicích k regresní přímce.

Regresní koeficient $b_3 = \operatorname{tg}[(\arctg 2 \alpha) : 2] \quad (2.69)$

$$\text{kde: } \operatorname{tg} 2\alpha = (2 \cdot S_{yx}) : (S_x^2 - S_y^2)$$

Konstanta posunutí $a_3 = \bar{y} - b_3 \cdot \bar{x} \quad (2.70)$

Reziduální směrodatná odchylka

$$s_{e3} = \sqrt{[(S_x^2 \cdot S_y^2) : (S_x^2 + S_y^2)] \cdot (1 - r_{yx}^2) \frac{n}{n-2}} \quad (2.71)$$

Směrodatná odchylka regresního koeficientu

$$s_{b3} = s_{e3} \cdot \sqrt{(\sin^4 \alpha) : (S_y^2 \cdot n)} \quad (2.72)$$

$$\text{kde: } \sin^4 \alpha = [\sin(\arctg b_3)]^4$$

Směrodatná odchylka konstanty posunutí

$$s_{a3} = s_{e3} \cdot \sqrt{x^2 : (S_x^2 \cdot n)} \quad (2.73)$$

2.9.1.5 Intervaly spolehlivosti regresní přímky

Meze spolehlivosti konfidenčního intervalu skutečné (neznámé) regresní přímky pro zvolenou hodnotu x_z jsou dány hranicemi

$$Y_z \pm t_\alpha(v) \cdot s_{Yz} = a + b \cdot x_z \pm t_\alpha(v) \cdot s_e \cdot \sqrt{\frac{1}{n} + [(x_z - \bar{x})^2 : (S_x^2 \cdot n)]} \quad (2.74)$$

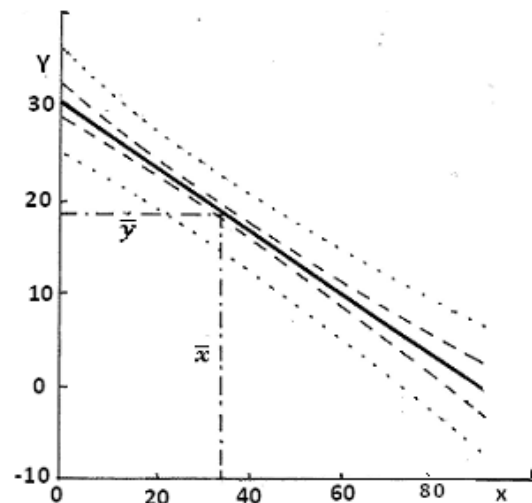
Při spojitě se měnících hodnotách x_z se též mění meze intervalu spolehlivosti a vytváří se pás kolem regresní přímky, jehož šíře je nejmenší v hodnotě $x_z = \bar{x}$. Vzdaluje-li se bod x_z od $\bar{x}$, šíře pásu monotónně roste (příklad obr. 2.9). Šíře pásu se zmenšuje, rostou-li hodnoty S_x^2 , n , tj. zvětšuje-li se variační rozpětí R_x a rozsah výběru.

Podobně jako směrodatná odchylka konstanty posunutí (s_a) je zvláštním případem směrodatné odchylky s_{YZ} , kdy $x_Z = 0$, je zvláštním případem i směrodatná odchylka rovná hodnotě $s_{\bar{Y}Z}$, kdy $x_Z = \bar{x}$ a lze ji odhadnout pomocí jednoduchého vztahu:

$$s_{\bar{Y}Z} = s_e : \sqrt{n} \quad (2.75)$$

V praxi nás častěji zajímají meze intervalu spolehlivosti, v nichž se mohou nacházet se zvolenou pravděpodobností $P(1-\alpha)$. 100% experimentální hodnoty mající souřadnice x_i, y_i , které představují možné výsledky měření na jednotce téhož základního souboru s danou pravděpodobností.

Na obr. 2.9 čárkované meze odpovídají mezím podle rovnice (2.74), predikce Y_i , zatímco tečkované odpovídají mezím podle rovnice (2.76), predikce experimentálních hodnot y_i



Obr.2.9: Pásy spolehlivosti regresní přímky

$$Y_Z \pm t_{\alpha}(v) \cdot s_{yZ} = a + b \cdot x_Z \pm t_{\alpha}(v) \cdot s_e \cdot \sqrt{\frac{1}{m} + \frac{1}{n} + [(x_Z - \bar{x})^2 : (S_x^2 \cdot n)]} \quad (2.76)$$

kde m je zvolený počet měření závisle proměnné y na jednom vzorku (obvykle $m = 1$ až 3), čl. 2.2.1, kdy $\bar{y}_m$ je průměr m měření hodnot y odpovídajících hodnotě x_Z .

V rovnicích 2.74, 2.76 stupně volnosti (v) = $(n - 2)$.

2.9.1.6 Praktické využití regresní přímky

2.9.1.6.1 Youdenova metoda identifikace druhů chyb (odchylek) [12]

K identifikaci druhů chyb (odchylek) zejména soustavných, při prověřování analytických metodik, laboratoří, přístrojů apod., je vhodná srovnávací metoda pomocí regresní přímky $Y = a_1 + b_1 \cdot x$ podle varianty $\sigma_x = 0, \sigma_y \neq 0$ (čl. 2.9.1.2). Prověřované hodnoty se použijí jako závisle proměnné a referenční hodnoty, u nichž je zajištěna prakticky nulová chyba (přesné množství standardní látky, etalonu známé, přesné hodnoty apod.), které tvoří párové referenční hodnoty x_i k prověřovaným hodnotám y_i .

Na obr. 2.9 A až E jsou schematicky zobrazeny vlivy různých druhů chyb. Ve všech případech plné úhlopříčky představují determinovaný vztah $Y = a + b \cdot x$, kde $a = 0, b = 1$. Naproti tomu vztahy, jejichž průběh je naznačen tečkováním a čárkováním, jsou výsledkem uplatnění soustavných chyb (odchylek) mimo obrázku A.

Obr. 2.9 A představuje praktickou shodnost prověřovaných výsledků s referenčními hodnotami $a_1 \approx 0, b_1 \approx 1$. Jedná se pouze o výskyt malých náhodných chyb, které lze ohodnotit pomocí reziduální směrodatné odchylky s_{e1} , případně pomocí směrodatných odchylek s_{a1}, s_{b1} nebo s_{yZ} (čl. 2.9.1.5). Protože statistiky t jsou též

$$t = (a - \underline{\alpha}) : s_a \quad (2.77)$$

$$t = (b - \underline{\beta}) : s_b \quad (2.78)$$

kde: $\underline{\alpha}$ = konstanta posunutí srovnávací přímky (= 0) - (úhlopříčka)
 $\underline{\beta}$ = regresní koeficient srovnávací přímky (= 1) - (úhlopříčka)

Ize t – testem otestovat hodnoty konstant a_1 , b_1 pomocí kritických hodnot $t_{\alpha}(v)$ tab. I., pro $(v) = (n - 2)$ a zvolenou pravděpodobnost $P(1-\alpha).100\%$.

Pokud bude $|a_1| : s_{a1} < t_{\alpha}(v)$ a $|b_1-1| : s_{b1} < t_{\alpha}(v)$, lze usuzovat, že se zvolenou pravděpodobností jsou prověřované hodnoty prosté vlivu soustavných chyb (odchylek) a naopak, jsou-li hodnoty $t_{\alpha}(v)$ menší než t statistiky, je se zvolenou pravděpodobností prověřovaný výběr ovlivněn soustavnými chybami (odchylkami).

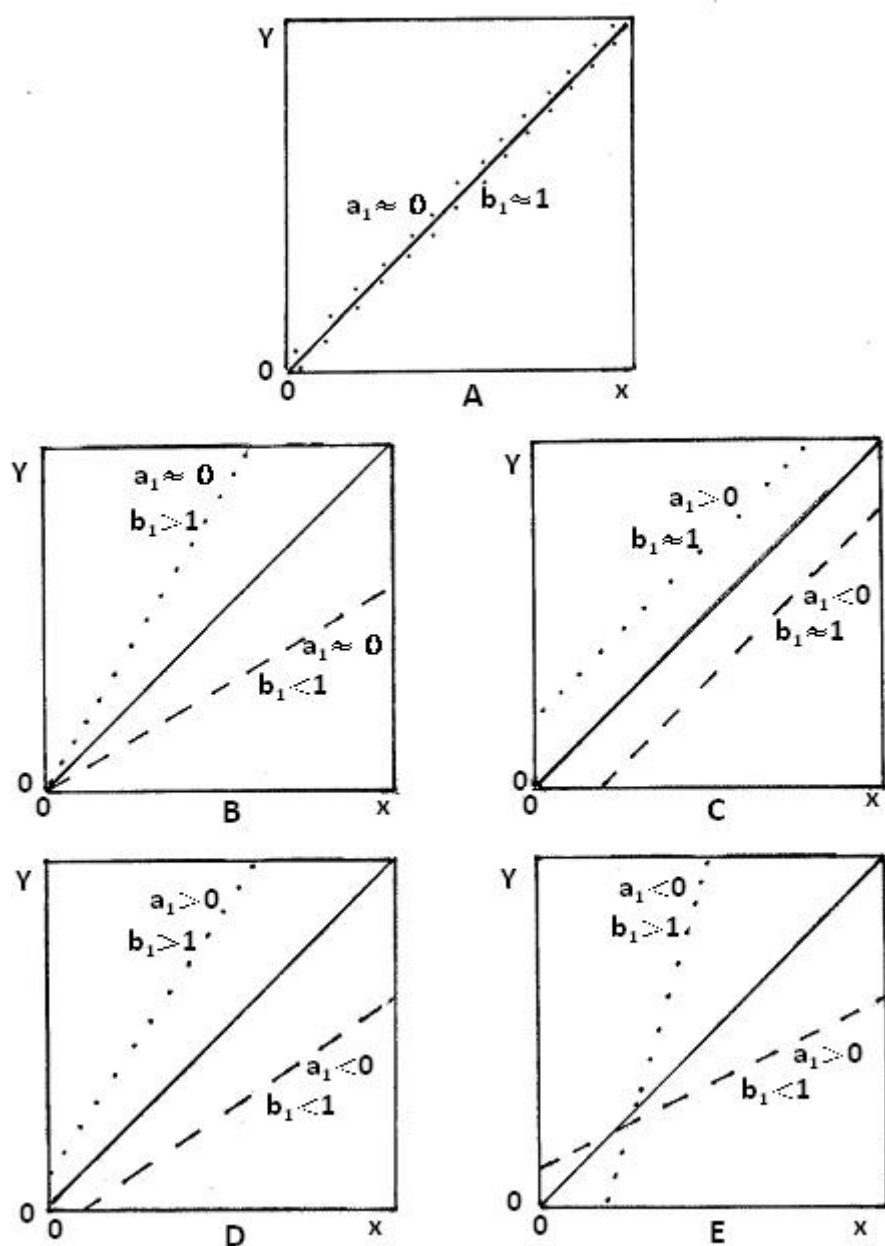
Obrázek 2.9 B znázorňuje závislosti ovlivněné přítomností soustavných proporcionálních chyb (odchylek), kdy $a_1 \approx 0$, $b_1 \neq 1$.

Obrázek C pak průběhem závislostí charakterizuje konstantní soustavné chyby (odchylky), kdy $a_1 \neq 0$, $b_1 \approx 1$ (závislosti jsou s uhlopříčkou).

Obrázky D a E vyznačují závislosti ovlivněné přítomností soustavných proporcionálních i konstantních chyb (odchylek), které se liší velikostí a směrem působení; při $a_1 \neq 0$, $b_1 \neq 1$ však zachovávají linearitu, kterou lze ověřit hodnotou korelačního koeficientu - viz čl. 2.9.1.1.

Ve všech reálných případech se však mohou vyskytovat chyby všech druhů. Hodnocení chyb pomocí Youdenovy metody lze použít úspěšně, je-li variabilita v hodnoceném variačním rozpětí konstantní. Chyby hrubé se testují na odlehlost podle čl. 2.9.1.5. V případě testování konstant a_1 , b_1 t-testem (2.77, 2.78), se test realizuje až po vyloučení vybočujících hodnot – čl. 2.9.1.7 a vypočtení nových parametrů regresní přímky $Y = a_1 + b_1 \cdot x$ bez vybočujících hodnot.

Chyby a odchylky mají rozdílnou genezi. Chyby nejsou zpravidla záměrné, zatímco odchylky mohou být výsledkem řízených podmínek pokusu a očekávány; za takových okolností je výraz chyba neadekvátní.



Obr.2.10 : Youdenova metoda identifikace druhů chyb

2.9.1.6.2 Tři varianty odhadu empirických regresních přímek

Zkušební příklad: Odhad parametrů lineární regresní závislosti spalného tepla bezvodého vzorku uhlí na obsahu popela v bezvodém vzorku uhlí. Spalné teplo Q_s^d je uvedeno v MJ/kg,

obsah popela A^d v hmotnostních %. Výpočet proveden pro všechny tři varianty. Pro kontrolu zkušebního výpočtu jsou uvedena všechna čísla z desetimístného displeje (fiktivní soubor).

Výběrový soubor

n	Q_s^d	A^d
	y	x
1	27,72	4,76
2	28,20	7,26
3	26,01	8,53
4	27,20	11,69
5	25,48	12,68
6	25,89	15,69
7	23,73	16,75
8	24,42	19,46
9	22,17	20,45
10	22,99	23,46
11	20,98	24,91
12	20,89	27,23
13	19,51	28,36
14	20,21	31,27
15	18,70	32,71
16	18,29	35,62
17	16,81	36,61
18	17,45	39,08
19	15,86	40,38
20	15,92	43,63
21	13,74	44,47
22	14,60	47,48
23	12,59	48,80
24	12,82	51,53
25	11,88	52,76
26	11,69	55,13
27	10,23	56,59
28	10,51	59,05
29	8,43	60,89
30	8,80	63,50

$\sum y = 553,72$
$\sum y^2 = 11287,1512$
$\sum x = 1020,73$
$\sum x^2 = 43724,8601$
$\sum yx = 15765,6105$
$\bar{y} = 18,45733333$
$\bar{x} = 34,02433333$
$\overline{x^2} = 1457,495337$
$S_y^2 = 35,56521956$
$S_x^2 = 299,8400779$
$S_{yx} = -102,4781118$
$r_{yx} = -0,992369771$

$\sigma_x = 0, \sigma_y \neq 0$
$b_1 = -0,341775897$
$a_1 = 30,0860304$
$s_{e1} = 0,761111207$
$s_{b1} = 8,024955893 \cdot 10^{-3}$
$s_{a1} = 0,306370003$
$\sigma_x \neq 0, \sigma_y = 0$
$b_2 = -0,347051862$
$a_2 = 30,26554158$
$s_{e2} = 2,209938616$
$s_{b2} = 0,06765608$
$s_{a2} = 0,889566315$
$\sigma_x \neq 0, \sigma_y \neq 0$
$b_3 = -0,342328526$
$a_3 = 30,10483322$
$s_{e3} = 0,71962791$
$s_{b3} = 2,310969552 \cdot 10^{-3}$
$s_{a3} = 0,289671732$

Z výsledků výpočtů je zřejmé, že nejnižší hodnoty parametrů variability poskytuje třetí varianta, kdy $\sigma_x \neq 0, \sigma_y \neq 0$ a MNČ je uplatněna na kolmicích vzdáleností experimentálních bodů od proložené regresní přímky. Regresní rovnice nejlepšího odhadu ze tří variant má podobu

$$Q_s^d = a_3 + b_3 \cdot A^d$$

kde: $a_3 = 30,10 \pm 0,29 \cdot t_{1-\alpha/2}(v)$,
 $b_3 = -0,3423 \pm 0,0023 \cdot t_{1-\alpha/2}(v)$, pro $(v) = n - 2$

2.9.1.7 Vylučování bodů odlehlých od regresní přímky

Odlehlost bodů bývá patrná již na vhodně upraveném bodovém grafu, v pravoúhlých kartézských souřadnicích, vyjadřujícím např. závislost $y = f(x)$, kde $x = A^d$, y vyjadřuje obsah organického prvku v bezvodé hmotě uhlí. Variabilita proměnných je v našich výběrech většinou způsobena přijatelnými místními změnami obsahu a složení buď anorganického, nebo organického podílu uhlí. Nápadně a značně odlehlé body od regresní přímky, na rozdíl od většiny přílehlajících bodů, bývají někdy výsledkem omylu, záměny proměnných, překlepu a evidentně do hodnoceného souboru nepatří. Do výběru však nepatří ani ty body, jejichž nápadná odlehlost je důsledkem zásadní změny chemického složení organického nebo anorganického podílu vzorku uhlí, tvořícího oddělené konkrce, případně samostatné vrstvy (např. pyrit, siderit, kalcit, křemen, fosilní pryskyřice apod.). Při vylučování odlehlých bodů použijeme kritéria $Y_z \pm t_{\alpha}(v) \cdot s_{yz}$ podle vztahu (2.76) a postupujeme tak, že: a) pro výběr odhadneme regresní rovnici $Y = a + b \cdot x$; b) hledáme body, jejichž difference δ_i od regresní přímky překračují meze intervalu spolehlivosti, tj.:

$$\delta_i = |y_i - Y_i| \geq s_{yz} \cdot t_{0,05}(v) \quad (2.79)$$

Přesnější kritérium v nerovnosti (2.79) lze nahradit přibližným kritériem $2,1 \cdot s_e$, zejména po transformaci nezávisle proměnné x na $1/x$. Existují-li hodnoty δ_i podle nerovnosti (2.79), nalezneme maximální hodnotu δ_c ; c) pokud hodnota $\delta_c = |y_c - Y_c|$, vyhovuje další nerovnosti:

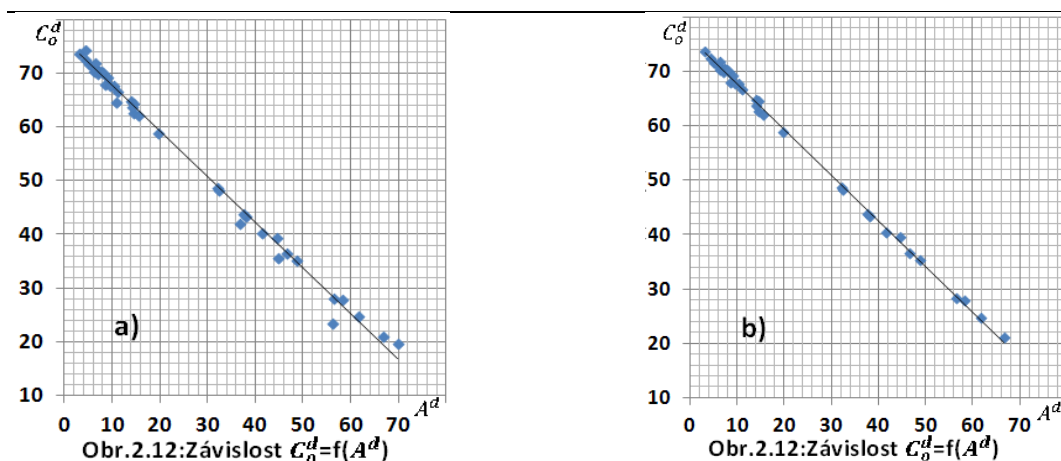
$$0,1 \geq \delta_c : \sum (y_i - Y_i)^2 \quad (2.80)$$

kde: $\sum (y_i - Y_i)^2$ = součet čtverců všech odchylek, pak se vypouštění odlehlých hodnot zastaví; d) pokud nevyhovuje hodnota δ_c nerovnosti (2.80), pak bod daný souřadnicemi (x_c, y_c) z výběru vyloučíme. Pro zbylý soubor o jednotku menší odhadneme opět regresní rovnici a pokračujeme od bodu b). Vypouštění odlehlých bodů tudíž postupuje po jednom bodu. Nerovnost (2.80) je odvozena z kritických hodnot Snedecorova rozdělení F.

Podmínkou je, že soubor zbavený hodnot zatížených nadlimitními odchylkami musí mít rozsah $n \geq n_{\min}$ a má mít co možno nejvyšší hodnotu variačního rozpětí R_x .

Došlo-li při vylučování odlehlých bodů ke zmenšení výběru tak, že podkročil rozsah $n_{\min}$ pro zvolenou P%, je nutné rozsah výběru zvětšit o experimentální data vyhovující požadavkům $n_{\min}$ a $|d_{\max}| \leq 2,1 \cdot s_e$. K výběru zbavenému odlehlých bodů přidáme další experimentální body v intervalu $(A_{\min}^d, A_{\max}^d)$, které mimo absence odlehlých hodnot vykazují také hodnoty nezávisle proměnné co nejvzdálenější od průměrné hodnoty $\bar{x}$. Cílem je dosažení maximálních hodnot R_x a S_x a tím i co nejvyššího koeficientu determinace ρ . Zvětšit rozsah výběru lze např. zmenšením úseků segmentovaného jádra vrtu.

Příklad způsobu vylučování odlehlých bodů je uveden pro závislost $C_o^d = f(A^d)$. Výběr vzorků uhlí z vrtného jádra z vrstvy cca 26m ze sokolovské pánve v rozsahu $n = 41$ byl analyzován a z hodnot obsahů organického uhlíku v bezvodém stavu C_o^d a obsahu popela v bezvodém stavu A^d byla vypočtena lineární regresní závislost. Na obrázcích 2.12 jsou grafy odpovídající původnímu výběru a) a zmenšenému výběru po vyloučení odlehlých bodů b) překračujících zvolený limit $\pm 1,96 \cdot s_e$. Vyloučeno 6 bodů. Zmenšený výběr splňoval požadavky $n \geq n_{\min}$. K výpočtu byla použita varianta podle čl. 2.9.1.2 a $u_p = 1,96$ pro dvoustranný kritický obor. Parametry regresní přímky z původního a zmenšeného výběru jsou uvedeny v tabulce 2.5.



Tabulka 2.5: Parametry regresní přímky z původního a zmenšeného výběru Obr. 2.12

Obr.	n	$\bar{x}$	$\bar{y}$	a	b	r_{yx}	R_x	S_x	s_e	s_a	$s_b \cdot 10^3$
a)	41	24,42	55,39	76,11	-0,848	-0,997	66,69	20,61	1,364	0,330	10,34
b)	35	22,25	57,44	76,12	-0,839	-0,999	63,51	19,29	0,649	0,168	5,692

Uvedený příklad představuje vylučování odlehlých bodů od regresní přímky, zatížených náhodnými chybami nadlimitní velikosti. V některých případech při použití modelu regresní přímky se však v reziduálním zbytku chyb může uplatnit další skrytá závislost proměnné y a to na hloubce uložení hodnocených vzorků v uhelné sloji. Jsou to zdánlivá hustota, která s rostoucí hloubkou též roste a obsah vody, který se z téhož důvodu zmenšuje se zmenšujícími se póry vlivem tlaku nadloží. Proto je někdy nutné prověřit, zda velká relativní chyba ($100 \cdot s_e : \bar{y}$) regresního odhadu modelem přímky není způsobena též závislostí na hloubce v m. Z jiných příčin to může být u obsahu dusíku, který v některých případech klesá směrem dolů od stropu sloje (pozorováno u některých průzkumných vrtů mosteckého a sokolovského uhlí).

2.9.2. Regresní závislost se dvěma nezávisle proměnnými – regresní rovina.

Podobně jako v čl. 2.9.1 se nejdříve vypočtou sumace proměnných, jejich druhých mocnin a součinů, dále jejich průměrné hodnoty až konečně variance a kovariance ($x_1 = A^d$, $x_2 = m$ hloubky), v přehledu takto:

- Sumace $\sum x_1, \sum x_2, \sum y, \sum x_1^2, \sum x_2^2, \sum y^2, \sum yx_1, \sum yx_2, \sum x_1x_2$,
- Průměry $\bar{x}_1, \bar{x}_2, \bar{y}, \bar{x}_1^2, \bar{x}_2^2, \bar{y}^2, \overline{yx_1}, \overline{yx_2}, \overline{x_1x_2}$,
- Variance $S_1^2 = \bar{x}_1^2 - \bar{x}_1^2, S_2^2 = \bar{x}_2^2 - \bar{x}_2^2, S_y^2 = \bar{y}^2 - \bar{y}^2$,
- Kovariance $S_{y1} = \overline{yx_1} - \bar{y} \cdot \bar{x}_1, S_{y2} = \overline{yx_2} - \bar{y} \cdot \bar{x}_2, S_{12} = \overline{x_1x_2} - \bar{x}_1 \cdot \bar{x}_2$.

Čísla v indexech pro nezávisle proměnné: 1 – rozumí se x_1 , 2 – rozumí se x_2 .

2.9.2.1. Korelační koeficienty

Párové korelační koeficienty

Má-li být prověřena těsnost korelačních vztahů proměnných pro regresní rovinu, pak v prvním kroku vypočteme párové korelační koeficienty.

$$r_{y1} = S_{y1} : \sqrt{S_1^2 \cdot S_y^2} \quad (2.83)$$

$$r_{y2} = S_{y2} : \sqrt{S_2^2 \cdot S_y^2} \quad (2.84)$$

$$r_{12} = S_{12} : \sqrt{S_1^2 \cdot S_2^2} \quad (2.85)$$

Pro párové korelační koeficienty se užívá kritických hodnot (tab. VII. s. 49) pro $v = n - 2$.

Dílčí korelační koeficienty

V dalším textu u závisle proměnné y na nezávisle proměnných x_1, x_2 charakterizují tečky mezi indexovým značením vztah mezi dvojicemi proměnných. Například koeficient $r_{y1.2}$ značí dílčí vztah mezi y a x_1 očištěný od vlivu x_2 a naopak $r_{y2.1}$ značí dílčí vztah mezi y a x_2 očištěný od vlivu x_1 . Těsnost závislosti x_1 na x_2 s vyloučením vlivu závisle proměnné y vyznačuje koeficient $r_{12.y}$.

Dílčí korelační koeficienty u dvojnásobné lineární regrese jsou dány vztahy

$$r_{y1.2} = (r_{y1} - r_{y2} \cdot r_{12}) : \sqrt{(1 - r_{y2}^2) \cdot (1 - r_{12}^2)} \quad (2.86)$$

$$r_{y2.1} = (r_{y2} - r_{y1} \cdot r_{12}) : \sqrt{(1 - r_{y1}^2) \cdot (1 - r_{12}^2)} \quad (2.87)$$

$$r_{12.y} = (r_{12} - r_{y1} \cdot r_{y2}) : \sqrt{(1 - r_{y1}^2) \cdot (1 - r_{y2}^2)} \quad (2.88)$$

Souhrnný korelační koeficient

Souhrnný korelační koeficient $r_{y.12}$ lze vypočítat také pomocí párových korelačních koeficientů. Jeho čtverec se podobně jako u párového korelačního koeficientu přímky nazývá koeficientem determinace a vynásoben stem $(100r_{y.12}^2)$ ukazuje, z kolika % lze změny závisle proměnné y vysvětlit zvolenou funkcí. U dílčích a souhrnného korelačního koeficientu se užívá kritických hodnot (tab. VII. s. 49) pro $v = n - 3$.

$$r_{y.12} = \sqrt{(r_{y1}^2 + r_{y2}^2 - 2 \cdot r_{y1} \cdot r_{y2} \cdot r_{12}) : (1 - r_{12}^2)} \quad (2.89)$$

U koeficientu $r_{y.12}$ tečka za y značí, že na hodnotu závisle proměnnou y působí obě nezávisle proměnné x_1, x_2 .

2.9.2.2. Koeficienty β

Koeficienty beta umožňují porovnávat intenzitu působení nezávisle proměnných x_1, x_2 na závisle proměnnou y bez ohledu na rozměry proměnných

$$\beta_{y1.2} = (r_{y1} - r_{y2} \cdot r_{12}) : (1 - r_{12}^2) \quad (2.90)$$

$$\beta_{y2.1} = (r_{y2} - r_{y1} \cdot r_{12}) : (1 - r_{12}^2) \quad (2.91)$$

2.9.2.3. Konstanty regresní rovnice roviny $Y = a_{y.12} + b_{y1.2} \cdot x_1 + b_{y2.1} \cdot x_2$

V následujícím textu jsou použity tečky mezi indexovým značením charakteristik: Např. $b_{y1.2}$ znamená, jak se změni závisle proměnná y při jednotkové změně x_1 při libovolné konstantní hodnotě x_2 za tečkou (vyloučíme x_2 z působení). Naopak $b_{y2.1}$ značí, jak se změni závisle proměnná y při jednotkové změně x_2 (před tečkou) při vyloučení působení x_1 (za tečkou). Tečka za y ($a_{y.12}$) znamená, že na hodnotu závisle proměnnou y působí obě nezávisle proměnné x_1, x_2 .

Výpočet regresních koeficientů je uveden dvěma způsoby: buď přímo, pomocí variancí a kovariancí, nebo, jsou-li k dispozici beta koeficienty, jednodušším postupem:

Regresní koeficienty

$$b_{y1.2} = (S_{y1} \cdot S_2^2 - S_{y2} \cdot S_{12}) : (S_1^2 \cdot S_2^2 - S_{12}^2) = (S_y \cdot \beta_{y1.2}) : S_1 \quad (2.92)$$

$$b_{y2.1} = (S_{y2} \cdot S_1^2 - S_{y1} \cdot S_{12}) : (S_1^2 \cdot S_2^2 - S_{12}^2) = (S_y \cdot \beta_{y2.1}) : S_2 \quad (2.93)$$

Aditivní konstanta posunutí

$$a_{y.12} = \bar{y} - b_{y1.2} \cdot \bar{x}_1 - b_{y2.1} \cdot \bar{x}_2 \quad (2.94)$$

2.9.2.4. Směrodatné odchyly parametrů regresní roviny

Reziduální směrodatná odchylna je dána vztahem

$$s_{y.12} = \sqrt{S_y^2 \cdot (1 - r_{y.12}^2) \cdot \frac{n}{n-3}} \quad (2.95)$$

Směrodatné odchyly regresních koeficientů

$$s_{by1.2} = s_{y.12} \cdot \sqrt{S_2^2 : (S_1^2 \cdot S_2^2 - S_{12}^2)} \quad (2.96)$$

$$s_{by2.1} = s_{y.12} \cdot \sqrt{S_1^2 : (S_1^2 \cdot S_2^2 - S_{12}^2)} \quad (2.97)$$

Směrodatná odchylna aditivní konstanty posunutí

$$s_{ay.12} = s_{y.12} \cdot \sqrt{\frac{1}{n} + [S_2^2 \cdot \bar{x}_1^2 + S_1^2 \cdot \bar{x}_2^2 - 2 \cdot S_{12} \cdot \bar{x}_1 \cdot \bar{x}_2] : (S_1^2 \cdot S_2^2 - S_{12}^2)} \quad (2.98)$$

Příklad ověření vícenásobné závislosti

Z vrtného jádra geologického průzkumu uhlí bylo odebráno 40 vzorků s vizuálně makro-petrografickým rozlišením. Byla hodnocena vrstva přibližně od stropu sloje, tj. cca od 116 až do 150m, tj. 0 až 34m. Po úpravě vzorků na zrno pod 0,2mm byly stanoveny obsahy vody, popela a dusíku a výsledky analýz přepočteny na bezvodý stav A^d, N^d . Pro výběr dat byla nejdříve odhadnuta regresní přímková závislost podle vztahu $N^d = f(A^d)$.

Parametry získané rovnice $Y = 1,049 - 0,0139 \cdot A^d$, $s_e = 0,149$, $r_{yx} = -0,83$, zvláště pak hodnoty relativní směrodatné odchylky $100s_e: \bar{y} \approx 19\%$ a koeficient determinace $\wp \approx 69\%$ naznačují, že rovnice regresní přímky pravděpodobně není ovlivněna pouze malými náhodnými chybami analytických stanovení znaků jakosti, ale bylo podezření na možnost vlivu další nezávisle proměnné, a to hloubky uložení odebraných vzorků a to od 0 do 34m délky jádra. Byly provedeny výpočty závislosti N^d na $x_1(A^d)$ a $x_2(m)$ s výsledky pro $n = 40$: korelační koeficienty: $r_{y1} = -0,83$; $r_{y2} = -0,7608$; $r_{12} = 0,3712$; $r_{y1.2} = -0,9087$; $r_{y2.1} = -0,8741$; $r_{y.12} = 0,9626$; beta koef. $\beta_{y1.2} = -0,6351$; $\beta_{y2.1} = -0,525$; přičemž relativní reziduální směrodatná odchylka $100s_{y.12}: \bar{y}$ činila 9,33% a koeficient determinace $\wp = 92,66\%$.

Výsledná regresní rovnice má podobu $N^d = 1,2402 - 0,01064 \cdot A^d - 0,01428 \cdot m$, kterou lze přibližně rozdělit na dvě rovnice:

$N_A^d = 0,947 - 0,01064 \cdot A^d$, která platí v intervalu 0 - 89% A^d , přičemž při $A^d = 89$ je $N_A^d = 0$.

$N_m^d = 0,293 - 0,01428 \cdot m$, která platí v intervalu 0 - 20,5m, přičemž při $m = 20,5$ je $N_m^d = 0$.

Výsledky regresního odhadu závislosti modelem přímky vedly k oprávněné predikci o pravděpodobném vlivu dvou nezávisle proměnných. Ukázalo se, že obsah dusíku v daném výběru vzorků z vrtného jádra je nepřímou závislostí v tomto případě nejen obligátně na obsahu popela, ale i na hloubce uložení uhlí ve sloji. Nejvyšší obsah dusíku pod stropem sloje se do hloubky snižuje až do cca 20,5m, kdy příspěvky dusíku z této závislosti k celkovému obsahu dusíku mizí. Závislost obsahu části dusíku v uhlí na hloubce uložení může být ovlivněna průsaky vodních srážek obsahujících sloučeniny dusíku z různých zdrojů.

2.10 Seznam tabulek I až VII

Tab.I: Kritické hodnoty „Studentova“ rozdělení t - str. 41

Hodnoty pro dvoustranný kritický obor pro $p = 1 - \alpha = 0,80; 0,90; 0,95; 0,98; 0,99; 0,999$; a $v = 1$ až 30 a dále 35; 40; 45; 60; 90; 200; 400; 700; 1000 v podobě čtyřmístných čísel. Jsou uvedeny i hodnoty p pro jednostranný kritický obor. Pro $v > 30$ platí přibližně [18]

$$t_p(v) \cong u_p \cdot \left[1 + \frac{1}{4v} (1 + u_p^2) + \frac{1}{96v^2} (3 + 16u_p^2 + 5u_p^4) \right] \quad (2.99)$$

kde: u_p značí kvantily normálního rozdělení pro $p = 1 - \alpha$.

Tab.II: Kvantily normálního rozdělení u_p - str. 42

II.a Hodnoty u_p pro dvoustranný kritický obor pro $p = 1 - \alpha = 0,8; 0,9; 0,95; 0,96; 0,975; 0,98; 0,99; 0,995; 0,996; 0,9975; 0,998; 0,999$; jsou uvedeny i hodnoty p pro jednostranný kritický obor. Kvantily u_p jsou sedmimístnými čísly.

II.b Hodnoty u_p pro dvoustranný kritický obor pro $p = 1 - \alpha$ od 0 do 0,99 po 0,01; od 0,99 do 0,999 po 0,001; od 0,999 do 0,9999 po 0,0001. Kvantily u_p pro $p > 0,7$ jsou čtyřmístnými čísly [1].

Tab.III: Kritické hodnoty Dean-Dixonova testu na odlehlost hodnot - str. 43

Třímístná čísla kritických hodnot pro $\alpha = 0,005; 0,010; 0,050; 0,100$; a pro $n = 3$ až 30.

Tab.IV: Kritické hodnoty Grubbsova testu na odlehlost hodnot - str. 44

Čtyřmístná čísla kritických hodnot pro $\alpha = 0,001; 0,005; 0,010; 0,025; 0,050; 0,100$; a pro $n = 3$ až 50. K aproximaci kritických hodnot pro $n > 50$ lze pro $\alpha = 0,05$ a $0,01$ použít vztahů [10]

$$G_{0,05,n} \cong 3,925 - 8,714 \cdot \left(\frac{1}{\sqrt{n}}\right) + 17,764 \cdot \left(\frac{1}{\sqrt{n}}\right)^2 - 24,198 \cdot \left(\frac{1}{\sqrt{n}}\right)^3 + 11,286 \cdot \left(\frac{1}{\sqrt{n}}\right)^4 \quad (2.100a)$$

$$G_{0,01,n} \cong 4,377 - 9,338 \cdot \left(\frac{1}{\sqrt{n}}\right) + 20,636 \cdot \left(\frac{1}{\sqrt{n}}\right)^2 - 38,248 \cdot \left(\frac{1}{\sqrt{n}}\right)^3 + 26,192 \cdot \left(\frac{1}{\sqrt{n}}\right)^4 \quad (2.100b)$$

Tab.V: Kritické hodnoty znaménkového testu - str. 45-46

Kritické hodnoty pro $\alpha = 0,01; 0,02; 0,05; 0,10; 0,20$; pro $n = 8$ až 500.

Tab.VI: Kvantily rozdělení χ^2 - str. 47-48

Hodnoty pro $1 - p = 0,005; 0,010; 0,025; 0,05$; a pro $p = 0,95; 0,975; 0,990; 0,995$; a pro stupně volnosti $v = 1$ až 90 ve tří až pětímístných číslech. Pro $v > 30$ lze pro výpočet hodnot χ^2 použít přibližných vztahů [13]

$$\chi_p^2(v) \cong v \cdot \left(1 - \frac{2}{9v} + u_p \cdot \sqrt{\frac{2}{9v}} \right)^3 \quad (2.101a)$$

$$\chi_{1-p}^2(v) \cong v \cdot \left(1 - \frac{2}{9v} - u_p \cdot \sqrt{\frac{2}{9v}} \right)^3 \quad (2.101b)$$

kde: u_p značí kvantily normálního rozdělení pro $p = 1 - \alpha$, viz tab. II., str. 42.

Při použití kvantilů rozdělení $\chi^2(v)$ z tabulky VI. na str. 47 se příslušné hodnoty χ^2 hledají pro stupně volnosti $(v) = (n - 1)$ a volí:

- a) při jednostranném kritickém oboru pro horní mez $1 - p$ a pro dolní mez p
- b) při oboustranném kritickém oboru se hodnoty χ^2 hledají při $(v) = (n - 1)$ stupních volnosti pro dolní mez (pro levou stranu nerovnosti) při $[(1 + p): 2]$ a pro horní mez (pro pravou stranu nerovnosti) při $\{(1 - p): 2\}$, přičemž $[(1 + p): 2] + [(1 - p): 2] = 1$.

Tab.VII: Kritické hodnoty korelačního koeficientu r_{yx} - str. 49

Hodnoty pro $\alpha = 0,10; 0,05; 0,02; 0,01; 0,001$; pro $v = n - 2 = 1$ až 100 [14]. Chybějící kritické hodnoty možno vypočítat podle vztahu

$$r_{\alpha}(v) = t_{\alpha}(v) : \sqrt{n - 2 + t_{\alpha}^2(v)} \quad [15] \quad (2.102)$$

kde: $t_{\alpha}(v)$ je kritická hodnota rozdělení t – viz tab. I. str. 41.

Tabulka I: Kritické hodnoty t_{α} rozdělení

$1-\alpha$	0,80	0,90	0,95	0,98	0,99	0,999	$1-\alpha$
α	0,20	0,10	0,05	0,02	0,01	0,001	α
v	(Dvoustranný kritický obor)						v
1	3,078	6,314	12,71	31,82	63,66	636,6	1
2	1,886	2,920	4,303	6,965	9,925	31,60	2
3	1,638	2,353	3,182	4,541	5,841	12,94	3
4	1,533	2,132	2,776	3,747	4,604	8,610	4
5	1,476	2,015	2,571	3,365	4,032	6,859	5
6	1,440	1,943	2,447	3,143	3,707	5,959	6
7	1,415	1,895	2,365	2,998	3,499	5,405	7
8	1,397	1,860	2,306	2,896	3,355	5,041	8
9	1,383	1,833	2,262	2,821	3,250	4,781	9
10	1,372	1,812	2,228	2,764	3,169	4,587	10
11	1,363	1,796	2,201	2,718	3,106	4,437	11
12	1,356	1,782	2,179	2,681	3,055	4,318	12
13	1,350	1,771	2,160	2,650	3,012	4,221	13
14	1,345	1,761	2,145	2,624	2,977	4,140	14
15	1,341	1,753	2,131	2,602	2,947	4,073	15
16	1,337	1,746	2,120	2,583	2,921	4,015	16
17	1,333	1,740	2,110	2,567	2,898	3,965	17
18	1,330	1,734	2,101	2,552	2,878	3,922	18
19	1,328	1,729	2,093	2,539	2,861	3,883	19
20	1,325	1,725	2,086	2,528	2,845	3,850	20
21	1,323	1,721	2,080	2,518	2,831	3,819	21
22	1,321	1,717	2,074	2,508	2,819	3,792	22
23	1,319	1,714	2,069	2,500	2,807	3,767	23
24	1,318	1,711	2,064	2,492	2,797	3,745	24
25	1,316	1,708	2,060	2,485	2,787	3,728	25
26	1,315	1,706	2,056	2,479	2,779	3,707	26
27	1,314	1,703	2,052	2,473	2,771	3,690	27
28	1,313	1,701	2,048	2,467	2,763	3,674	28
29	1,311	1,699	2,045	2,462	2,756	3,659	29
30	1,310	1,697	2,042	2,457	2,750	3,646	30
35	1,306	1,690	2,030	2,438	2,724	3,591	35
40	1,303	1,684	2,021	2,423	2,704	3,551	40
45	1,301	1,679	2,014	2,412	2,690	3,520	45
60	1,296	1,671	2,000	2,390	2,660	3,460	60
90	1,291	1,662	1,987	2,369	2,632	3,402	90
200	1,286	1,653	1,972	2,345	2,600	3,340	200
400	1,284	1,649	1,966	2,336	2,588	3,315	400
700	1,283	1,647	1,963	2,332	2,583	3,305	700
1000	1,282	1,646	1,962	2,330	2,581	3,300	1000
∞	1,282	1,645	1,960	2,326	2,576	3,291	∞
v	(Jednostranný kritický obor)						v
α	0,10	0,05	0,025	0,01	0,005	0,0005	α
$1-\alpha$	0,90	0,95	0,975	0,99	0,995	0,9995	$1-\alpha$

Tabulky IIa,b: Kvantily u_p normálního rozdělení

Tabulka IIa.: Kvantily u_p normálního rozdělení

↓ Dvoustranný kritický obor ↓							
p_2	0,8	0,9	0,95	0,96	0,975	0,98	p_2
u_p	1,281552	1,644854	1,959964	2,053749	2,241403	2,326348	u_p
p_1	0,9	0,95	0,975	0,98	0,9875	0,99	p_1
↑ Jednostranný kritický obor ↑							
$p = 1 - \alpha$							
↓ Dvoustranný kritický obor ↓							
p_2	0,99	0,995	0,996	0,9975	0,998	0,999	p_2
u_p	2,575829	2,807034	2,878162	3,023335	3,090232	3,290527	u_p
p_1	0,995	0,9975	0,998	0,99875	0,999	0,9995	p_1
↑ Jednostranný kritický obor ↑							

Tabulka IIb.: Kvantily u_p normálního rozdělení

Dvoustranný kritický obor										
$1 - \alpha =$ $= p_2$	0	1	2	3	4	5	6	7	8	9
	u_p									
0,0	0,000	013	025	038	050	063	075	088	100	113
0,1	0,126	138	151	164	176	189	202	215	228	240
0,2	0,253	266	279	293	305	319	332	345	358	372
0,3	0,385	399	412	426	440	454	468	482	496	510
0,4	0,524	539	553	568	583	598	613	628	643	659
0,5	0,674	690	706	722	739	755	772	789	806	824
0,6	0,842	860	878	896	915	935	954	974	994	1,015
0,7	1,036	058	080	103	126	150	175	200	227	254
0,8	1,282	311	341	372	405	440	476	514	555	598
0,9	1,645	695	751	812	881	1,960	2,054	2,170	2,326	2,576
0,99	2,576	612	652	697	748	807	878	2,968	3,090	3,291
0,999	3,291	320	353	390	432	481	540	615	719	891

Tabulka III: Kritické hodnoty Dean-Dixonova testu

n	α				n
	0,005	0,010	0,050	0,100	
3	0,994	0,988	0,941	0,886	3
4	0,926	0,889	0,765	0,679	4
5	0,821	0,780	0,642	0,557	5
6	0,740	0,698	0,560	0,482	6
7	0,680	0,637	0,507	0,434	7
8	0,634	0,590	0,468	0,399	8
9	0,598	0,555	0,437	0,370	9
10	0,568	0,527	0,412	0,349	10
11	0,542	0,502	0,392	0,332	11
12	0,522	0,482	0,376	0,318	12
13	0,503	0,465	0,361	0,305	13
14	0,488	0,450	0,349	0,294	14
15	0,475	0,438	0,338	0,285	15
16	0,463	0,426	0,329	0,277	16
17	0,452	0,416	0,320	0,269	17
18	0,442	0,407	0,313	0,263	18
19	0,433	0,398	0,306	0,258	19
20	0,425	0,391	0,300	0,252	20
21	0,418	0,384	0,295	0,247	21
22	0,411	0,378	0,290	0,242	22
23	0,404	0,372	0,285	0,238	23
24	0,399	0,367	0,281	0,234	24
25	0,393	0,362	0,277	0,230	25
26	0,388	0,357	0,273	0,227	26
27	0,384	0,353	0,269	0,224	27
28	0,380	0,349	0,266	0,220	28
29	0,376	0,345	0,263	0,218	29
30	0,372	0,341	0,260	0,215	30

Tabulka IV: Kritické hodnoty Grubbsova testu

n	α						n
	0,001	0,005	0,010	0,025	0,050	0,100	
3	1,414	1,414	1,414	1,414	1,412	1,406	3
4	1,731	1,728	1,723	1,710	1,689	1,645	4
5	1,990	1,972	1,955	1,917	1,869	1,791	5
6	2,203	2,160	2,130	2,067	1,996	1,894	6
7	2,377	2,310	2,265	2,082	2,093	1,974	7
8	2,521	2,431	2,374	2,273	2,172	2,041	8
9	2,643	2,532	2,464	2,349	2,238	2,097	9
10	2,747	2,616	2,540	2,414	2,294	2,146	10
11	2,837	2,689	2,606	2,470	2,343	2,190	11
12	2,915	2,753	2,663	2,519	2,387	2,229	12
13	2,984	2,809	2,713	2,563	2,426	2,264	13
14	3,046	2,859	2,759	2,602	2,461	2,297	14
15	3,102	2,905	2,800	2,638	2,494	2,327	15
16	3,152	2,946	2,837	2,670	2,523	2,354	16
17	3,198	2,983	2,871	2,701	2,551	2,380	17
18	3,240	3,017	2,903	2,728	2,577	2,404	18
19	3,278	3,049	2,932	2,754	2,601	2,426	19
20	3,314	3,079	2,959	2,779	2,623	2,447	20
21	3,347	3,106	2,984	2,801	2,644	2,467	21
22	3,378	3,132	3,008	2,823	2,664	2,486	22
23	3,407	3,156	3,030	2,843	2,683	2,504	23
24	3,434	3,179	3,051	2,862	2,701	2,521	24
25	3,459	3,200	3,071	2,880	2,718	2,537	25
26	3,483	3,220	3,089	2,897	2,734	2,553	26
27	3,506	3,239	3,107	2,913	2,749	2,568	27
28	3,528	3,258	3,124	2,929	2,764	2,582	28
29	3,548	3,275	3,140	2,944	2,778	2,596	29
30	3,567	3,291	3,156	2,958	2,792	2,609	30
31	3,586	3,307	3,171	2,972	2,805	2,622	31
32	3,603	3,322	3,185	2,985	2,818	2,634	32
33	3,620	3,337	3,199	2,998	2,830	2,646	33
34	3,636	3,351	3,212	3,010	2,842	2,657	34
35	3,652	3,364	3,224	3,022	2,853	2,668	35
36	3,667	3,377	3,236	3,033	2,864	2,679	36
37	3,681	3,389	3,248	3,044	2,874	2,689	37
38	3,695	3,401	3,259	3,055	2,885	2,699	38
39	3,708	3,413	3,270	3,065	2,894	2,709	39
40	3,720	3,424	3,281	3,075	2,904	2,718	40
41	3,733	3,435	3,291	3,084	2,913	2,727	41
42	3,745	3,445	3,301	3,094	2,922	2,736	42
43	3,756	3,455	3,310	3,103	2,931	2,745	43
44	3,767	3,465	3,320	3,112	2,940	2,753	44
45	3,778	3,474	3,329	3,120	2,948	2,762	45
46	3,788	3,483	3,338	3,129	2,956	2,770	46
47	3,798	3,492	3,346	3,137	2,964	2,778	47
48	3,808	3,501	3,354	3,145	2,972	2,785	48
49	3,818	3,510	3,363	3,152	2,980	2,793	49
50	3,827	3,518	3,370	3,160	2,987	2,800	50

Tabulka V: Kritické hodnoty znaménkového testu

n	$\alpha=0,01$		$\alpha=0,02$		$\alpha=0,05$		$\alpha=0,10$		$\alpha=0,20$		n
	k_1	k_2	k_1	k_2	k_1	k_2	k_1	k_2	k_1	k_2	
8	0	8	0	8	0	8	1	7	1	7	8
9	0	9	0	9	1	8	1	8	2	7	9
10	0	10	0	10	1	9	1	9	2	8	10
11	0	10	1	10	1	10	2	9	2	9	11
12	1	11	1	11	2	10	2	10	3	9	12
13	1	12	1	12	2	11	3	10	3	10	13
14	1	13	2	12	2	12	3	11	4	10	14
15	2	13	2	13	3	12	3	12	4	11	15
16	2	14	2	14	3	13	4	12	4	12	16
17	2	15	3	14	4	13	4	13	5	12	17
18	3	15	3	15	4	14	5	13	5	13	18
19	3	16	4	15	4	15	5	14	6	13	19
20	3	17	4	16	5	15	5	15	6	14	20
21	4	17	4	17	5	16	6	15	7	14	21
22	4	18	5	17	5	17	6	16	7	15	22
23	4	19	5	18	6	17	7	16	7	16	23
24	5	19	5	19	6	18	7	17	8	16	24
25	5	20	6	19	7	18	7	18	8	17	25
26	6	20	6	20	7	19	8	18	9	17	26
27	6	21	7	20	7	20	8	19	9	18	27
28	6	22	7	21	8	20	9	19	10	18	28
29	7	22	7	22	8	21	9	20	10	19	29
30	7	23	8	22	9	21	10	20	10	20	30
31	7	24	8	23	9	22	10	21	11	20	31
32	8	24	8	24	9	23	10	22	11	21	32
33	8	25	9	24	10	23	11	22	12	21	33
34	9	25	9	25	10	24	11	23	12	22	34
35	9	26	10	25	11	24	12	23	13	22	35
36	9	27	10	26	11	25	12	24	13	23	36
37	10	27	10	27	12	25	13	24	14	23	37
38	10	28	11	27	12	26	13	25	14	24	38
39	11	28	11	28	12	27	13	26	14	25	39
40	11	29	12	28	13	27	14	26	15	25	40
41	11	30	12	29	13	28	14	27	15	26	41
42	12	30	13	29	14	28	15	27	16	26	42
43	12	31	13	30	14	29	15	28	16	27	43
44	13	31	13	31	15	29	16	28	17	27	44
45	13	32	14	31	15	30	16	29	17	28	45
46	13	33	14	32	15	31	16	30	18	28	46
47	14	33	15	32	16	31	17	30	18	29	47
48	14	34	15	33	16	32	17	31	19	29	48
49	15	34	15	34	17	32	18	31	19	30	49
50	15	35	16	34	17	33	18	32	19	31	50
52	16	36	17	35	18	34	19	33	20	32	52
54	17	37	18	36	19	35	20	34	21	33	54
56	17	39	18	38	20	36	21	35	22	34	56
58	18	40	19	39	21	37	22	36	23	35	58
60	19	41	20	40	21	39	23	37	24	36	60
62	20	42	21	41	22	40	24	38	25	37	62

Tabulka V. Kritické hodnoty znaménkového testu- pokr.

n	$\alpha=0,01$		$\alpha=0,02$		$\alpha=0,05$		$\alpha=0,10$		$\alpha=0,20$		n
	k_1	k_2	k_1	k_2	k_1	k_2	k_1	k_2	k_1	k_2	
62	20	42	21	41	22	40	24	38	25	37	62
64	21	43	22	42	23	41	24	40	26	38	64
66	22	44	23	43	24	42	25	41	27	39	66
68	22	46	23	45	25	43	26	42	28	40	68
70	23	47	24	46	26	44	27	43	29	41	70
72	24	48	25	47	27	45	28	44	30	42	72
74	25	49	26	48	28	46	29	45	30	44	74
76	26	50	27	49	28	48	30	46	31	45	76
78	27	51	28	50	29	49	31	47	32	46	78
80	28	52	29	51	30	50	32	48	33	47	80
82	28	54	30	52	31	51	33	49	34	48	82
84	29	55	30	54	32	52	33	51	35	49	84
86	30	56	31	55	33	53	34	52	36	50	86
88	31	57	32	56	34	54	35	53	37	51	88
90	32	58	33	57	35	55	36	54	38	52	90
92	33	59	34	58	36	56	37	55	39	53	92
94	34	60	35	59	37	57	38	56	40	54	94
96	34	62	36	60	37	59	39	57	41	55	96
98	35	63	37	61	38	60	40	58	42	56	98
100	36	64	37	63	39	61	41	59	43	57	100
110	41	69	42	68	44	66	45	65	47	63	110
120	45	75	46	74	48	72	50	70	52	68	120
130	49	81	51	79	53	77	55	75	57	73	130
140	54	86	55	85	57	83	59	81	61	79	140
150	58	92	60	90	62	88	64	86	66	84	150
160	63	97	64	96	67	93	69	91	71	89	160
170	67	103	69	101	71	99	73	97	76	94	170
180	72	108	73	107	76	104	78	102	80	100	180
190	76	114	78	112	81	109	83	107	85	105	190
200	81	119	83	117	85	115	87	113	90	110	200
220	90	130	92	128	94	126	97	123	99	121	220
240	99	141	101	139	104	136	106	134	103	131	240
260	108	152	110	150	113	147	116	144	119	141	260
280	117	163	120	160	123	157	125	155	128	152	280
300	127	173	129	171	132	168	135	165	138	162	300
320	136	184	138	182	141	179	144	176	148	172	320
340	145	195	148	192	151	189	154	186	157	183	340
360	155	205	157	203	160	200	163	197	167	193	360
380	164	216	166	214	170	210	173	207	177	203	380
400	173	227	176	224	179	221	183	217	186	214	400
420	183	237	185	235	189	231	192	228	196	224	420
440	192	248	195	245	198	242	202	238	206	234	440
460	201	259	204	256	208	252	211	249	215	245	460
480	211	269	214	266	218	262	221	259	225	255	480
500	220	280	223	277	227	273	231	269	235	265	500

Tabulka VI: Kvantily rozdělení $\chi^2(v)$

v	1-p				p				v
	0,005	0,010	0,025	0,050	0,950	0,975	0,990	0,995	
1	—	—	0,001	0,004	3,842	5,024	6,635	7,879	1
2	0,010	0,020	0,051	0,103	5,992	7,378	9,210	10,597	2
3	0,072	0,115	0,216	0,352	7,815	9,348	11,345	12,838	3
4	0,207	0,297	0,484	0,711	9,488	11,143	13,277	14,860	4
5	0,412	0,554	0,831	1,146	11,070	12,833	15,086	16,750	5
6	0,676	0,872	1,237	1,635	12,592	14,449	16,812	18,548	6
7	0,989	1,239	1,690	2,167	14,067	16,013	18,475	20,278	7
8	1,344	1,647	2,180	2,733	15,507	17,535	20,090	21,955	8
9	1,735	2,088	2,700	3,325	16,919	19,023	21,666	23,589	9
10	2,156	2,558	3,247	3,940	18,307	20,483	23,209	25,188	10
11	2,603	3,054	3,816	4,575	19,675	21,920	24,725	26,757	11
12	3,074	3,571	4,404	5,226	21,026	23,337	26,217	28,300	12
13	3,565	4,107	5,009	5,892	22,362	24,736	27,688	29,819	13
14	4,075	4,660	5,629	6,571	23,685	26,119	29,141	31,319	14
15	4,601	5,229	6,262	7,261	24,996	27,488	30,578	32,891	15
16	5,142	5,812	6,908	7,962	26,296	28,845	32,000	34,267	16
17	5,697	6,408	7,564	8,672	27,587	30,191	33,409	35,718	17
18	6,265	7,015	8,231	9,391	28,869	31,526	34,805	37,156	18
19	6,844	7,633	8,907	10,117	30,144	32,852	36,191	38,582	19
20	7,434	8,260	9,591	10,851	31,410	34,170	37,566	39,997	20
21	8,034	8,897	10,283	11,591	32,671	35,479	38,932	41,401	21
22	8,643	9,543	10,982	12,338	33,924	36,781	40,289	42,796	22
23	9,260	10,196	11,689	13,091	35,172	38,076	41,638	44,181	23
24	9,886	10,856	12,401	13,848	36,415	39,364	42,980	45,559	24
25	10,520	11,524	13,120	14,611	37,652	40,646	44,314	46,928	25
26	11,160	12,198	13,844	15,379	38,885	41,923	45,642	48,290	26
27	11,808	12,879	14,573	16,151	40,113	43,195	46,963	49,645	27
28	12,461	13,565	15,308	16,928	41,337	44,461	48,278	50,993	28
29	13,121	14,256	16,047	17,708	42,557	45,722	49,588	52,336	29
30	13,787	14,953	16,791	18,493	43,773	46,979	50,892	53,672	30
31	14,458	15,655	17,539	19,281	44,985	48,232	52,191	55,003	31
32	15,134	16,362	18,291	20,072	46,194	49,480	53,486	56,328	32
33	15,815	17,074	19,047	20,867	47,400	50,725	54,776	57,648	33
34	16,501	17,789	19,806	21,664	48,602	51,966	56,061	58,964	34
35	17,192	18,509	20,569	22,465	49,802	53,203	57,342	60,275	35
36	17,887	19,233	21,336	23,269	50,998	54,437	58,619	61,581	36
37	18,586	19,960	22,106	24,075	52,192	55,668	59,892	62,883	37
38	19,289	20,691	22,878	24,884	53,384	56,896	61,162	64,181	38
39	19,996	21,426	23,654	25,695	54,572	58,120	62,428	65,476	39
40	20,707	22,164	24,433	26,509	55,758	59,342	63,691	66,766	40
41	21,421	22,906	25,215	27,326	56,942	60,561	64,950	68,053	41
42	22,138	23,650	25,999	28,144	58,124	61,777	66,206	69,336	42
43	22,859	24,398	26,785	28,965	59,304	62,990	67,459	70,616	43
44	23,584	25,148	27,575	29,787	60,481	64,201	68,710	71,893	44
45	24,311	25,901	28,366	30,612	61,656	65,410	69,957	73,166	45

Tabulka VI.: Kvantily rozdělení $\chi^2(v)$ pokračování

v	1-p				p				v
	0,005	0,010	0,025	0,050	0,950	0,975	0,990	0,995	
46	25,041	26,657	29,160	31,439	62,830	66,617	71,201	74,437	46
47	25,775	27,416	29,956	32,268	64,001	67,821	72,443	75,704	47
48	26,511	28,177	30,755	33,098	65,171	69,023	73,683	76,969	48
49	27,249	28,941	31,555	33,930	66,339	70,222	74,919	78,231	49
50	27,991	29,707	32,357	34,764	67,505	71,420	76,154	79,490	50
51	28,735	30,475	33,162	35,600	68,669	72,616	77,386	80,747	51
52	29,481	31,246	33,968	36,437	69,832	73,810	78,616	82,001	52
53	30,230	32,018	34,776	37,276	70,993	75,002	79,843	83,253	53
54	30,981	32,793	35,586	38,116	72,153	76,192	81,069	84,502	54
55	31,735	33,570	36,398	38,958	73,311	77,380	82,292	85,749	55
56	32,490	34,350	37,212	39,801	74,468	78,567	83,513	86,994	56
57	33,248	35,131	38,027	40,646	75,624	79,752	84,733	88,236	57
58	34,008	35,913	38,844	41,492	76,778	80,936	85,950	89,477	58
59	34,770	36,698	39,662	42,339	77,931	82,117	87,166	90,715	59
60	35,534	37,485	40,482	43,188	79,082	83,298	88,379	91,952	60
61	36,301	38,273	41,303	44,038	80,232	84,476	89,591	93,186	61
62	37,068	39,063	42,126	44,889	81,381	85,654	90,802	94,419	62
63	37,838	39,855	42,950	45,741	82,529	86,830	92,010	95,649	63
64	38,610	40,649	43,776	46,595	83,675	88,004	93,217	96,878	64
65	39,383	41,444	44,603	47,450	84,821	88,177	94,422	98,105	65
66	40,158	42,240	45,431	48,305	85,965	90,349	95,626	99,330	66
67	40,935	43,038	46,260	49,162	87,108	91,519	96,828	100,55	67
68	41,713	43,838	47,092	50,020	88,250	92,689	98,028	101,78	68
69	42,494	44,639	47,924	50,879	89,391	93,856	99,228	103,00	69
70	43,275	45,442	48,758	51,739	90,531	95,023	100,43	104,21	70
71	44,058	46,246	49,592	52,600	91,670	96,189	101,62	105,43	71
72	44,843	47,051	50,428	53,462	92,808	97,353	102,82	106,65	72
73	45,629	47,858	51,265	54,325	93,945	98,516	104,01	107,86	73
74	46,417	48,666	52,103	55,189	95,081	99,678	105,20	109,07	74
75	47,206	49,475	52,942	56,054	96,217	100,84	106,39	110,29	75
76	47,997	50,286	53,782	56,920	97,351	102,00	107,58	111,50	76
77	48,788	51,097	54,623	57,786	98,484	103,16	108,77	112,70	77
78	49,582	51,910	55,466	58,654	99,617	104,32	109,96	113,91	78
79	50,376	52,725	56,309	59,522	100,75	105,47	111,14	115,12	79
80	51,172	53,540	57,153	60,391	101,88	106,63	112,33	116,32	80
81	51,969	54,357	57,998	61,261	103,01	107,78	113,51	117,52	81
82	52,767	55,174	58,845	62,132	104,14	108,94	114,69	118,73	82
83	53,567	55,993	59,692	63,004	105,27	110,09	115,88	119,93	83
84	54,368	56,813	60,540	63,876	106,39	111,24	117,06	121,13	84
85	55,170	57,634	61,389	64,749	107,52	112,39	118,24	122,32	85
86	55,973	58,456	62,239	65,623	108,65	113,54	119,41	123,52	86
87	56,777	59,279	63,089	66,498	109,77	114,69	120,59	124,72	87
88	57,582	60,103	63,941	67,373	110,90	115,84	121,77	125,91	88
89	58,389	60,928	64,793	68,249	112,02	116,99	122,94	127,11	89
90	59,196	61,754	65,647	69,126	113,15	118,14	124,12	128,30	90

Tabulka VII: Kritické korelační koeficienty r_{α}

ν n-2	α					ν n-2
	0,10	0,05	0,02	0,01	0,001	
1	0,98769	0,99692	0,999507	0,999877	0,9999988	1
2	0,90000	0,95000	0,98000	0,990000	0,99900	2
3	0,8054	0,8783	0,93433	0,95873	0,99116	3
4	0,7293	0,8114	0,8822	0,91720	0,97406	4
5	0,6694	0,7545	0,8329	0,8745	0,95074	5
6	0,6215	0,7067	0,7887	0,8343	0,92493	6
7	0,5822	0,6664	0,7498	0,7977	0,8982	7
8	0,5494	0,6319	0,7155	0,7646	0,8721	8
9	0,5214	0,6021	0,6851	0,7348	0,8471	9
10	0,4973	0,5760	0,6581	0,7079	0,8233	10
11	0,4762	0,5529	0,6339	0,6835	0,8010	11
12	0,4575	0,5324	0,6120	0,6614	0,7800	12
13	0,4409	0,5139	0,5923	0,6411	0,7603	13
14	0,4259	0,4973	0,5742	0,6226	0,7420	14
15	0,4124	0,4821	0,5577	0,6055	0,7246	15
16	0,4000	0,4683	0,5425	0,5897	0,7084	16
17	0,3887	0,4555	0,5285	0,5751	0,6932	17
18	0,3783	0,4438	0,5155	0,5614	0,6787	18
19	0,3687	0,4329	0,5034	0,5487	0,6652	19
20	0,3598	0,4227	0,4921	0,5368	0,6524	20
21	0,3515	0,4132	0,4815	0,5256	0,6402	21
22	0,3438	0,4044	0,4716	0,5151	0,6287	22
23	0,3365	0,3961	0,4622	0,5052	0,6178	23
24	0,3297	0,3882	0,4534	0,4958	0,6073	24
25	0,3233	0,3809	0,4451	0,4869	0,5974	25
26	0,3172	0,3739	0,4372	0,4785	0,5880	26
27	0,3115	0,3673	0,4297	0,4705	0,5791	27
28	0,3061	0,3610	0,4226	0,4629	0,5703	28
29	0,3009	0,3550	0,4157	0,4556	0,5620	29
30	0,2960	0,3494	0,4093	0,4487	0,5541	30
35	0,2746	0,3246	0,3810	0,4182	0,5189	31
40	0,2573	0,3044	0,3578	0,3932	0,4896	40
45	0,2428	0,2875	0,3284	0,3721	0,4648	45
50	0,2306	0,2732	0,3218	0,3541	0,4433	50
55	0,2201	0,2609	0,3074	0,3385	0,4244	55
60	0,2108	0,2500	0,2948	0,3248	0,4078	60
70	0,1954	0,2319	0,2737	0,3017	0,3799	70
80	0,1829	0,2172	0,2565	0,2830	0,3568	80
90	0,1726	0,2050	0,2422	0,2673	0,3375	90
100	0,1638	0,1946	0,2301	0,2540	0,3211	100

Literatura 2. kapitoly

- [1] BÖHM, J.: Vyrovnávací počet, SNTL Praha 1964.
- [2] RESSLER, J. – RIEDL, R.: Sborník VŠCHT v Praze, D 24, Technologie paliv, SPN Praha 1971.
- [3] ČSN 01 5110: Vzorkování materiálů. Základní ustanovení. 1974.
- [4] ČSN 01 5111: Vzorkování sypkých a zrnitých materiálů. 1974.
- [5] ECKSCHLAGER, K.: Chyby chemických rozborů. SNTL Praha 1971.
- [6] ČERMÁK, V.: Výběrové statistické zjišťování. SNTL Praha 1980.
- [7] DEAN, R.B. – DIXON, W.J.: Anal. Chem. 23, 636, 1951.
- [8] GRUBBS, F.E.: Sample Criteria for Testing Outlying Observations. Annals of Mathematical Statistics 21,1950,s.27-58.
- [9] ECKSCHLAGER, K.- HORSÁK, I.- KODEJŠ, Z.: Vyhodnocování analytických výsledků a metod. SNTL Praha 1980.
- [10] DOERFFEL, K.- ECKSCHLAGER, K.: Optimální postupy chemické analýzy. SNTL Praha 1985.
- [11] ANDĚL, J.: Matematická statistika. SNTL Praha 1978.
- [12] YODEN, W.J.: Statistical Methods for Chemists. Wiley, New York 1951.
- [13] ČSN 01 0223: Pravidla stanovení odhadů a konfidenčních mezí. 1990.
- [14] ČSN 01 0252: Statistické metody v průmyslové praxi II.- korelace a regrese. 1976
- [15] MÜLLER, P.H.-NEUMANN, P.-STORM, R.: Tafeln der mathematischen Statistik. VEB Fachbuchverlag Leipzig DDR 1973
- [16] JÍLEK, M.: Statistické toleranční meze. SNTL Praha 1988, 280 s.
- [17] MELOUN, M.- MILITKÝ, J.: Statistická analýza experimentálních dat. ACADEMIA 2004.
- [18] LIKEŠ, J.-LAGA, J.: Základní statistické tabulky. SNTL Praha, 1978.
- [19] KUBANT, J.: Teoretické základy hodnocení složení a vlastností uhlí. UHLÍ, 34, s.406, 1986.

3. HODNOCENÍ ANALÝZ TUHÝCH PALIV

3.1 Fosilní tuhá paliva

Přibližně od konce 19. století je známo, že je rozdíl mezi obsahem popelovin v uhlí a množstvím popela vzniklého po spálení uhlí. Od té doby do současnosti bylo v řadě prací na celém světě vynaloženo nemalé úsilí na vypracování metodiky stanovení popelovin, v nich vázané hydrátové vody a popelového faktoru. Zároveň bylo zřejmé, že stanovené či vypočtené hodnoty popelovin a dalších charakteristik nemají reprezentovat pouze jeden určitý vzorek, ale mají představovat průměrnou hodnotu v množině vzorků odebraných například z průzkumného vrtu, záseku, vrstvy nebo určité lokality. Zprvu se předpokládalo, že popelový faktor je konstantou blízká se hodnotě 1,1. Řada autorů např. [1, 2, 3] navrhla k odhadu obsahu popelovin různé vzorce, jejichž přesnost je problematická zejména s ohledem na podíl hydrátové vody (W_M) [4, 5], jejíž obsah je uzančnými vzorci velmi nepřesně odhadnut. Poměr $W_M: Al_2O_3$ se např. podle experimentálních výsledků [6] pouze u kaolinitů pohybuje v rozpětí 0,33 - 0,38. Za přítomnosti dalších jílových minerálů, např. dickitu, halloysitu, montmorillonitu, illitu a bentonitu, se může tento poměr nalézat v rozmezí až 0,2 - 0,6, což vylučuje možnost použití nějakých konstantních poměrů. Bez znalosti složení minerálů umožňuje ČSN 44 1364 [7] metodou regresní analýzy odhadnout obsah hydrátové vody, odpovídající průměru ve volných popelovinách, určité vybrané množiny vzorků uhlí daného vrtu, záseku nebo lokality.

K experimentálnímu stanovení obsahu popelovin byla vypracována řada metod: dlouhodobá nízkoteplotní oxidace podle Browna a kol. [8], kyselinová digesce s HCl a HF podle Radmachera a Mohrhauera [9], nízkoteplotní oxidace excitovaným kyslíkem ve VF poli podle Gleita a Hollanda [10], kvantitativní planimetrická optická analýza [11] a konečně grafická extrapolační stanovení [12, 13]. Žádný z uvedených postupů nezaručuje stanovení správných hodnot obsahu popelovin, ať již proto, že při stanovení dochází k nepředpokládaným reakcím (při spalovacích a digesčních metodách), nebo z důvodů metodických chyb (při grafické extrapolaci) či subjektivních chyb (při mikroskopických postupech). Některé metody jsou příliš časově náročné nebo náročné na přístrojové vybavení nebo kvalifikaci pracovníka; žádná z nich však nedává spolehlivé výsledky.

Podobně je tomu při nepřímém experimentálním stanovení obsahu hydrátové vody založeném na předpokladu, že minerály obsažené v uhlí mají stejné složení jako nadložní či podložní horniny provázející uhelné sloje [14, 15]: tento předpoklad nemusí být splněn.

3.1.1 Dosavadní přepočty na hořlavinu

Rozdíl mezi obsahem popelovin a popela je důsledkem termického oxidačního rozkladu anorganických minerálů, sloučenin organické hmoty uhlí s anorganickými prvky a chemických reakcí mezi rozkladnými oxidačními produkty. Velikost rozdílů bývá (zejména u méně prouhelných uhlí) ovlivněna ve větší míře únikem hydrátové vody z jílových minerálů, dále rozkladem karbonátů železa a vápníku, disulfidů železa, oxidací sloučenin železa, reakcemi oxidů síry s alkalickými oxidy a řadou dalších reakcí, např. oxidací prvků vázaných jako humáty apod. V dalším textu označování analytických ukazatelů tuhých paliv viz ČSN 44 1310 [16]. Při přepočtech výsledků analýz na tzv. hořlavinu – rozumí se bezvodý a bezpopelný stav – například podle vzorce

$$Y^{daf} = (100 \cdot Y^d) : (100 - A^d) \quad (3.1)$$

se neuvažují hmotnostní změny při přeměně popelovin na popel, což vede k chybám, a to tím větším, čím větší je obsah popela. Správný výsledek se získá podle vzorce

$$Y_o^o = (100 \cdot Y^d) : (100 - A^d \cdot f_A) \quad (3.2)$$

kde: $A^d \cdot f_A = M^d$ = obsah popelovin v bezvodém stavu, v hmotnostních %,

f_A = popelový faktor,

Y_o^o = znak jakosti v organické hmotě (teoretická hodnota),

a to pouze tehdy, pokud Y^d při A_{max}^d má nulovou hodnotu Y_o^d , jak bude vysvětleno v čl. 3.1.4.

K jak velkým chybám dochází při přepočtech znaků jakosti na hořlavinu při použití vzorce (3.1) je patrné z tabulky 3.1, zatímco správné hodnoty odpovídající správné organické hmotě jsou označeny Y_o^o a jsou stejné pro jakýkoliv výchozí obsah popela v intervalu (A_{min}^d, A_{max}^d). Označení o v indexu znamená, že hodnota Y_o^d je při A_{max}^d nulová. Výběr vzorků byl získán odběrem z vrtného jádra hnědého uhlí SHD, který po úpravě vzorků na zrno pod 0,2mm, stanovení obsahů C,H,V, přepočtu na bezvodý stav, vyjádření jejich lineární závislosti na obsahu popela rovnicemi typu $Y^d = a + b \cdot A^d$ a po odstranění odlehlých hodnot, dosáhl rozsahu $n = 60$. Tento výběr označený n60 bude dále v příkladech několikrát použit. Z regresních rovnic pak byly vypočteny hodnoty odpovídající zvoleným obsahům A^d a též jim odpovídající hodnoty popelového faktoru.

Tabulka 3.1: Přepočty podle vzorců (3.1) a (3.2)

A^d	5	10	20	40	60	80
f_A	0,8483	0,9911	1,0626	1,0983	1,1102	1,11617
C_o^o	73,27	73,27	73,27	73,27	73,27	73,27
C_o^{daf}	73,86	73,34	72,12	68,47	61,16	39,23
H_o^o	5,40	5,40	5,40	5,40	5,40	5,40
H_o^{daf}	5,44	5,40	5,31	5,04	4,50	2,89
H_t^{daf}	5,54	5,61	5,77	6,27	7,27	10,25
V_o^o	48,76	48,76	48,76	48,76	48,76	48,76
V^{daf}	49,86	50,32	51,40	54,65	61,14	80,62

Z ukázky chyb, jichž se dopouštíme použitím vzorce (3.1) při přepočtech na „hořlavinu“, je zřejmé, že problém vyžaduje konečné řešení, zejména při hodnocení vysokopopelnatých paliv, a to i s důsledky pro revizi příslušných norem.

3.1.2 Teoretické principy pracovní hypotézy

Z tabulky 3.1 je zřejmé, že popelový faktor f_A není konstantou, ale je výsledkem působení dvou faktorů podle vztahu vyjadřujícího vážený průměr

$$f_A = (f_q \cdot q^d + f_p \cdot p^d) : (q^d + p^d) \quad (3.3)$$

$$\text{přičemž } q^d + p^d = A^d \quad (3.4)$$

kde: A^d = obsah směsného popela v bezvodém stavu uhlí, v hmotn. %,

q^d = obsah popela odpovídající volným popelovinám v bezvodém stavu uhlí, v hmotn. %,

p^d = obsah popela odpovídající vázaným popelovinám v bezvodém stavu uhlí, v hmotn. %,

f_q = popelový faktor pro popel volných popelovin (konstantní),
 f_p = popelový faktor pro popel vázaných popelovin (konstantní).

Existence vázaných a volných popelovin je axiomem pracovní hypotézy formulované již dříve [20], která vychází z modelu současného složení uhlí a nevyžaduje znalost geneze a původu popelovin. Hlavní charakteristikou vázaných popelovin je, že při různém obsahu veškerých popelovin vázané popeloviny doprovázejí organickou hmotu ve stálém hmotnostním poměru a zachovávají stálý obsah jednotlivých prvků, které tvoří vázané popeloviny. Tento konstantní poměr popelovin k organické hmotě a složení může být pro různá uhlí, popř. i pro různé vrstvy uhlí téže sloje různý a tím charakteristický pro hodnocenou vrstvu.

Vázanými popelovinami mohou být popeloviny vázané chemicky na organickou uhelnou hmotu (např. humáty vápníku, železa, hliníku apod.), prvky vázané v některých minerálech, např. Whewellit (oxalátu vápenatém), Humboldtinu (oxalátu železa), Mellitu (hlinité soli kyseliny mellitové-benzenhexakarbonové) apod. Vazba prvků na organickou hmotu uhlí může mít jak iontový, tak chelátový charakter. Přibližně lze vázané popeloviny považovat za ten anorganický podíl uhlí, který je běžným způsobem mechanicky neoddělitelný. Není rozhodující, zda mají vázané popeloviny svůj původ ve vlastních popelovinách rostlin, z nichž uhlí vzniklo, či zda se do uhlí dostaly v kterémkoliv stadiu geneze uhelné sloje. Hmotnost vázaných popelovin je obvykle nižší než z nich vzniklý popel zejména s navázaným kyslíkem, sírany, karbonáty, halogeny a dalšími prvky a oxidy.

Hlavní charakteristikou volných popelovin je, že se ve hmotě uhlí mohou vyskytovat v množství od 0 do 100% hmotnosti, což znamená, že mohou vytvářet i samostatné vrstvy.

Podstatnou složkou volných popelovin jsou obvykle fylosilikáty ze skupin kaolinitů + jílových montmorillonitů + slíd, jejichž obsah se zvětšuje s obsahem popela. Dalšími minerály tvořícími volné popeloviny, někdy i významnějšího zastoupení, jsou karbonáty vápníku, hořčíku a železa, dále disulfidy železa, minerály oxidu křemičitého, popř. menší množství dalších minerálů, jejichž přehledy podávají např. publikace [21 – 26]. Významný úbytek hmotnosti volných popelovin při spalování uhlí činí hydrátová voda.

Přibližně lze volné popeloviny považovat za ten anorganický podíl uhlí, který je mechanicky oddělitelný. Je-li v analyzovaném vzorku 100% volných popelovin, pak není přítomna organická hmota a nejsou ani přítomny vázané popeloviny.

Hodnoty A^d , p^d , q^d se mohou vyskytovat pouze v reálném intervalu (A_{min}^d, A_{max}^d) , kde

A_{min}^d = minimální obsah popela v bezvodém uhlí, v hmotnostních %,

A_{max}^d = maximální obsah popela v bezvodém uhlí, v hmotnostních %.

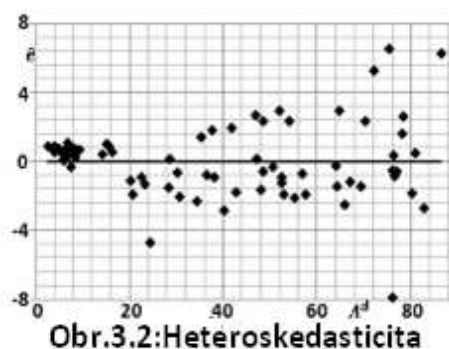
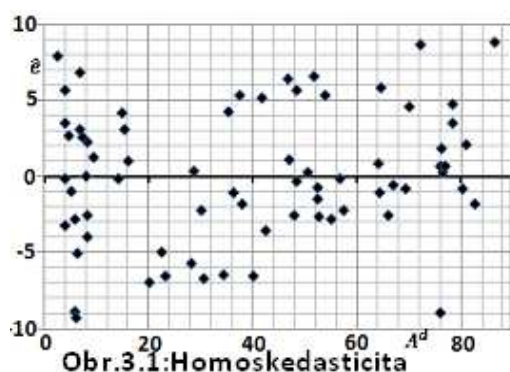
Minimální obsah popela je nejnižší obsah popela vypočtený, tj. odhadnutý pro základní soubor pomocí zpracovaného výběru z „homogenní vrstvy uhlí“. Tato hodnota představuje obsah popela, který bychom stanovili spálením vzorku uhlí neobsahujícího volné popeloviny, ale pouze vázané. Maximální obsah popela je nejvyšší možný obsah vypočtený, tj. odhadnutý pro základní soubor pomocí zpracovaného výběru z „homogenní vrstvy uhlí“.

„Homogenní vrstvou uhlí“ se rozumí taková modelová vrstva, která má konstantní poměr vázaných popelovin k organické hmotě uhlí a konstantní složení organické hmoty i vázaných a volných popelovin. Je přirozené, že ve skutečnosti budou tyto hodnoty na rozdíl od teoretického modelu více či méně variabilní. To bývá způsobeno výskytem vrstev či konkrétních některých minerálů, např. pyritu, sideritu, kalcitu, křemene, případně organických látek, např. vosků, pryskyřic apod. Tyto extrémy se občas vyskytnou ve složení vzorků a také bývají vyloučeny jako

odlehle hodnoty od regresní závislosti sledovaného znaku na obsahu popela. To však nebrání použitelnosti této pracovní hypotézy, na níž je založena metoda hodnocení analýz uhlí pomocí lineárních regresních závislostí znaků jakosti uhlí na obsahu popela. Navíc lze přesnost odhadů vypočítat. Extrémní odchylky od regresních závislostí vykazují často korelaci u znaků Q_s^d, H^d, T_K^d, V^d a $1/d_r^d$, přinejmenším v souhlasných znaménkách.

3.1.3 Odhady hodnot základních charakteristik výběru

Popsané postupy se poněkud liší od ČSN 44 1364 [7], protože jak od původních prací, které byly podklady [20,27] pro tvorbu ČSN [7], tak od jejího vydání uplynula řada let a postupy doznaly jistého vývoje. K výpočtu se použije způsob regrese podle varianty v čl. 2.9.1.2. Pro regresní odhady závislostí obsahů oxidů v popelu na obsahu popela byla ve studii [20] jako nejvhodnější lineární model zvolena rovnice $OX_i^A = a + b \cdot (1:A^d)$ (3.4). Důvod byl ten, že zde nedochází k transformaci závisle proměnné včetně reziduí $\hat{e}_i = (y_i - Y_i)$. V ČSN 44 1364 však byl použit nevhodný lineární model $SiO_2^d = f(A^d)$ (3.5), u něhož dochází k transformaci závisle proměnné, včetně reziduí, jak je patrné z rovnice $SiO_2^d = SiO_2^A \cdot (A^d:100)$. Rozdíly zmíněných modelů jsou zřejmé ze vztahu reziduí k nezávisle proměnné (A^d) na obrázcích 3.1 a 3.2.



Obrázek 3.1 představuje „mrak“ reziduí přibližně náhodného rozdělení u vhodného modelu pro použití MNČ. Na obrázku 3.2, kdy došlo k transformaci závisle proměnné a reziduí, dostal „mrak“ reziduí podobu výseče, což nasvědčuje heteroskedasticitě závisle proměnné, včetně reziduí $(y - Y)$.

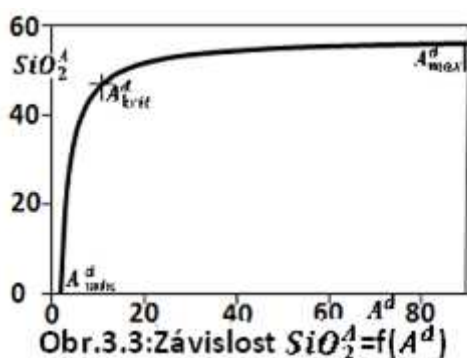
Tato skutečnost vyžaduje revizi ČSN 44 1364 [7], a proto v této práci používáme model regresní rovnice $OX_i^A = a_i + b_i \cdot (1:A_j^d)$ (3.4), kde OX_i^A značí obsah i-tého oxidu v popelu, v hmotnostních % závislý na reciproké hodnotě j-tého obsahu popela, v hmotnostních %. Heteroskedasticita závisle proměnné při použití MNČ vede ke zkreslení odhadů konstant regresní rovnice. Ke konstrukci obrázků 3.1 a 3.2 byl použit výběr označený n75 s větší variabilitou reziduí $\hat{e}$ pro názornost významu volby modelu regresní rovnice.

Základní informace pro odhad hodnot $A_{max}^d, f_q, A_{min}^d, f_p, A_{krit}^d, f_A$ se pro konkrétní výběr získají z regresních rovnic:

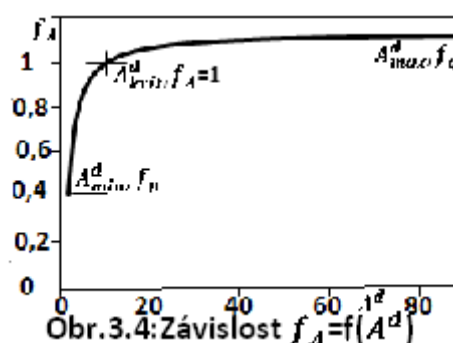
- a) Závislosti obsahu uhlíku v organické hmotě bezvodého uhlí na obsahu popela v bezvodém uhlí

$$C_t^d - C_M^d = C_o^d \quad (3.6) \quad C_o^d = a_c + b_c \cdot A^d \quad (3.7) \quad A_{max}^d = |a_c: b_c| \quad (3.8) \quad f_q = 100: A_{max}^d \quad (3.9)$$

- b) Závislosti obsahu oxidu křemičitého v popelu na reciproké hodnotě obsahu popela v bezvodém uhlí. Byla použita reciproká hodnota A^d a tím závislost linearizována.



Obr. 3.3: Závislost $SiO_2^A = f(A^d)$



Obr. 3.4: Závislost $f_A = f(A^d)$

$$SiO_2^A = a_{SiO_2} + b_{SiO_2} \cdot (1:A^d) \quad (3.10)$$

$$A_{min}^d = |b_{SiO_2} : a_{SiO_2}| \quad (3.11)$$

$$„A“ \text{ v exponentu značí obsah v popelu. } A_{krit}^d = \sqrt{|b_{SiO_2}|} \quad (3.12)$$

Interpretace průběhu závislosti oxidu křemičitého na popelu a situování hodnoty A_{krit}^d je však názornější podle funkce, kde je nezávisle proměnnou A^d , jejímž grafickým znázorněním je hyperbola a hodnotou nezávisle proměnné ve vrcholu této hyperboly je právě A_{krit}^d .

Kritický obsah popela A_{krit}^d představuje dělicí mez, od níž napravo, se zvyšujícím se obsahem popela, převažuje vliv volných popelovin. Naopak, od této dělicí hodnoty nalevo, se snižujícím se obsahem popela, převažuje vliv vázaných popelovin, viz obr. 3.3. Důležité zároveň je, že při hodnotě A_{krit}^d odpovídá funkce hodnot $f_A = 1$, viz obr. 3.4, a to i při hodnotě $(1:A_{krit}^d)$ v rovnici

$$f_A = a_k + b_k \cdot (1:A_{krit}^d) = 1$$

- c) Dalším krokem je sestavení linearizované rovnice závislosti popelového faktoru f_A na reciproké hodnotě obsahu popela

$$f_A = a_k + b_k \cdot (1:A^d) \quad (3.13)$$

pomocí konstant z následujících rovnic

$$b_k = (f_q - 1) : [(1:A_{max}^d) - (1:A_{krit}^d)] \quad (3.14)$$

$$a_k = 1 - b_k \cdot (1:A_{krit}^d) \quad (3.15)$$

- d) Předposlední základní charakteristiku odhadneme pomocí rovnice (3.13) dosazením hodnoty

$$(1:A_{min}^d); \quad f_p = a_k + b_k \cdot (1:A_{min}^d) \quad (3.16)$$

- e) Poslední charakteristiku odhadneme postupnou transformací rovnice (3.13); nejdříve vynásobením obou stran hodnotou A^d dostaneme

$$f_A \cdot A^d = M^d = a_k \cdot A^d + b_k \quad (3.17)$$

a poté odečtením obou stran rovnice (3.17) od 100 dostaneme

$$100 - M^d = 100 - b_k - a_k \cdot A^d = \text{závislost organické hmoty uhlí na obsahu popela} \quad (3.18)$$

Po úpravě rovnice (3.18) a novém označení dostaneme

$$\Theta_o^d = a_\Theta + b_\Theta \cdot A^d \quad (3.19)$$

kde: Θ_o^d = obsah organické hmoty v bezvodém uhlí, v hmotnostních %,

a_Θ = konstanta posunutí = $100 - b_k$

b_Θ = regresní koeficient = $-a_k$

Vydělením $100:a_\Theta$ získáme faktor f_a , který umožňuje odhad hodnot Y_o^o pouhým součinem

$$Y_o^o = a_Y \cdot f_a \quad (3.20)$$

kde: Y_o^o = znak jakosti v organické hmotě uhlí,

a_Y = konstanta posunutí závislosti libovolného prvku organické hmoty na obsahu popela,

např. (a_C, a_H, a_N, a_S, a_O).

$$f_a = 100:a_\Theta \quad (3.21)$$

Příklad odhadu hlavních (zaokrouhlených) charakteristik výběru n60

V příkladu byl použit výběr n60, a to již při zhotovení tabulky 3.1 a obrázků 3.3 a 3.4.

$$C_o^d = 74,317 - 0,8309 \cdot A^d \quad SiO_2^A = 57,292 - 113,66 \cdot (1:A^d)$$

$$A_{max}^d = 89,44157 \quad A_{krit}^d = 10,66114 \quad A_{min}^d = 1,98387 \quad f_q = 1,11805 \quad f_p = 0,41379$$

$$f_A = 1,13402 - 1,42884 \cdot (1:A^d) \quad \Theta_o^d = 101,42884 - 1,13402 \cdot A^d \quad f_a = 0,98591$$

$$M_{max}^d = 100 = a_k \cdot A_{max}^d + b_k = A_{max}^d \cdot f_q \quad M_{min}^d = 0,82091 = a_k \cdot A_{min}^d + b_k = A_{min}^d \cdot f_p$$

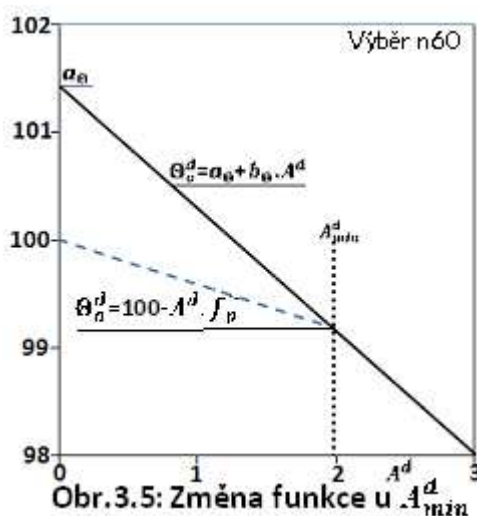
Při numerických výpočtech je vhodné používat všech číslic nejméně desetimístního displeje.

3.1.4 Vlastnosti námi formulovaných regresních závislostí

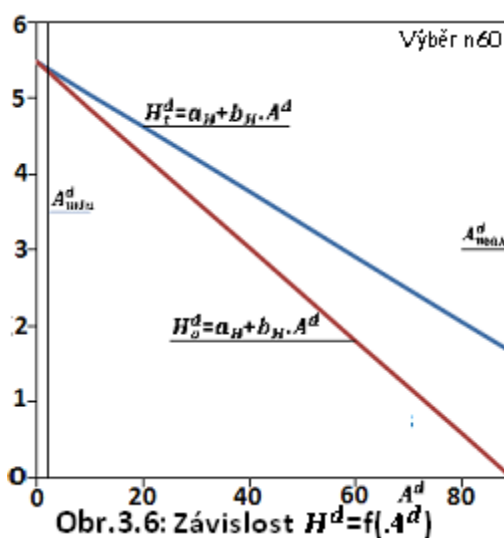
Závislost obsahu uhlíku na obsahu popela zpravidla poskytuje nejpřesnější regresní závislosti, snadno se koriguje na obsah karbonátového uhlíku, je prvkem s nejvyšším obsahem v organické hmotě, regresní přímka protíná osu x v nejvzdálenějším místě od nuly a tak určuje místo hodnoty obsahu A_{max}^d a snadného výpočtu faktoru f_q , a konečně poskytuje jeden ze dvou bodů pro konstrukci linearizované závislosti pro výpočet f_A . Závislost obsahu oxidu křemičitého na reciproké hodnotě obsahu popela byla zvolena proto, že obsah SiO_2 je zpravidla nejvyšším obsahem oxidu ve volných popelovinách, takže jeho nulový obsah indikuje nepřítomnost volných popelovin a tím zároveň přítomnost pouze vázaných popelovin a minimální obsah popela. Tato závislost zároveň umožňuje jednoduchý odhad obsahu popela A_{krit}^d , při němž nabývá faktor f_A hodnoty 1 a tím poskytuje druhý bod pro konstrukci rovnice pro výpočet faktoru f_A . Lineární závislost organické hmoty v bezvodém uhlí na obsahu popela se při snižování obsahu popela od A_{max}^d po A_{min}^d nemění. Avšak při teoretickém snížení pod A_{min}^d se markantně změní průběh této základní funkce, jak je znázorněno na obr. 3.5 čárkovaně přímkou směřující ku 100.

Původní přímka směřuje k hodnotě a_θ . Pod hodnotou A_{min}^d se faktor f_A závislý na A^d mění v konstantní faktor pro vázané popeloviny f_p . Podobné změny lze očekávat u všech znaků jakosti, které jsou buď složkou organické hmoty, nebo charakterizují její vlastnosti. Výpočet znaku jakosti uhlí v organické hmotě podle vzorce (3.2) je správný pouze tehdy, je-li znak jakosti zcela samostatný, tj. očištěný od jiných, doprovodných složek, jako například $C_t^d - C_M^d = C_o^d$ nebo $H_t^d - H_M^d = H_o^d$. Index o značí organickou hmotu, která je při A_{max}^d nulová. To je případ rovnice $H_o^d = a_H + b_H \cdot A^d$ na obr. 3.6.

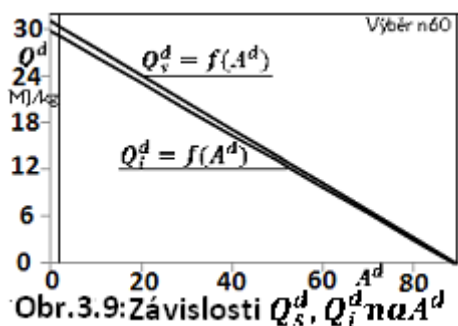
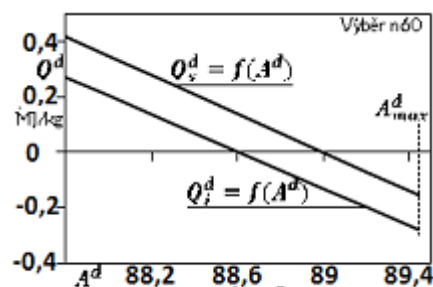
Poté rovnice $H_t^d = a_H + b_H \cdot A^d$ při A_{max}^d vykazuje hodnotu H_M , což odpovídá hydrátové vodě volných popelovin W_M^d bez analytické vody odstraněné při 105 - 110°C. Obě rovnice mají stejnou konstantu posunutí a_H . Úprava regresního koeficientu horní rovnice je snadná: $b_{Ho} = -a_H : A_{max}^d$. Upravený koeficient se společnou konstantou posunutí pak lze použít při výpočtu H_o^d podle vzorce (3.2) při jakémkoliv obsahu popela z intervalu (A_{min}^d, A_{max}^d) . Snadnějším postupem je však součin $H_o^d = a_H \cdot f_A$.



Obr. 3.5: Změna funkce u A_{min}^d



Obr. 3.6: Závislost $H^d = f(A^d)$

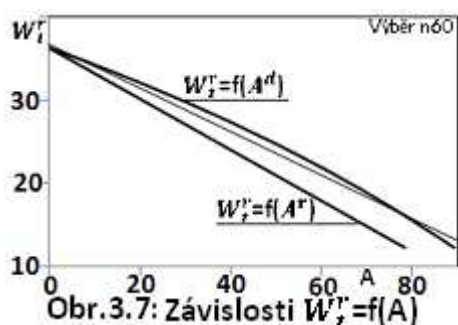
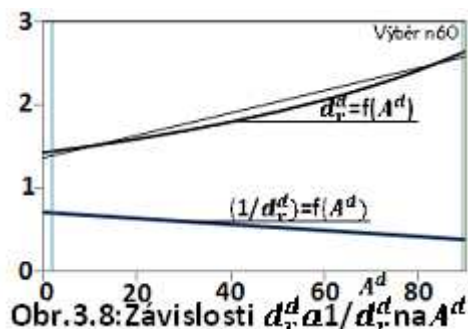
Obr. 3.9: Závislosti Q_s^d, Q_i^d na A^d 

Obr. 3.10: Výšeč Obr. 3.9

Vlastností konstant posunutí závislostí prvků organické hmoty na obsahu popela je, že $a_\Theta = (a_C + a_H + a_N + a_S + a_O)$, resp., že $100 = a_\Theta \cdot f_a$. To umožňuje odhad závislosti např. obsahu kyslíku na obsahu popela $O_o^d = a_o + b_o \cdot A^d$, kde $a_o = a_\Theta - (a_C + a_H + a_S + a_N)$ a $b_o = -a_o / A_{max}^d$.

Podmínkou úspěšného hodnocení závislostí je dodržení linearit funkcí. Prvními příklady byly linearizace rovnic (3.10) a (3.13). Dalším příkladem je dodržování stejného stavu [16] proměnných, tj. např. závisle i nezávisle proměnná je v původním stavu a grafická podoba závislosti na obr. 3.7 je evidentně lineární. Naproti tomu, je-li závisle proměnná v původním stavu a nezávisle proměnná v bezvodém stavu, je průběh funkce nelineární, jak je patrné z horní křivky, jejíž konkávní zakřivení je zvýrazněno proloženou přímkou. Linearity lze dosáhnout též úpravou závisle proměnné na reciprokovou hodnotu, jako je tomu např. u závislosti reciprokové hodnoty skutečné hustoty bezvodého uhlí na obsahu popela v bezvodém uhlí, jak je znázorněno na obr. 3.8. Konkávní zakřivení je též zvýrazněno proložením křivky přímkou. Zaznamenáváme-li skutečnou hustotu v g/cm^3 na tři desetinná místa [28], je vhodné použít při regresních výpočtech reciprokových hodnot minimálně se čtyřmi desetinnými místy, neboť reciproková hodnota hustoty je někdy až téměř o řád menší.

Při spalování vzorku uhlí v kalorimetrické bombě dochází mimo spálení organické hmoty též k oxidační, tepelně zabarvené rozkladné reakci popelovin – celková závislost na obsahu popela viz obr. 3.9. Výsledný tepelný efekt exo- a endotermických reakcí je možné odhadnout i pro volné popeloviny dosazením hodnoty obsahu A_{max}^d do rovnic závislostí spalného tepla a výhřevnosti na obsahu popela.

Obr. 3.7: Závislosti W_s^r, W_i^r na A Obr. 3.8: Závislosti $d_r^d, 1/d_r^d$ na A^d

Například pro výběr n60 byly výsledné termické efekty $Q_{sM}^d = -0,153 \text{ MJ/kg}$ a pro $Q_{iM}^d = -0,278 \text{ MJ/kg}$. Pro malé hodnoty bylo grafické znázornění možné pouze z výše celkové závislosti na obr. 3.10, kde jsou výsledné termické efekty v průsečících s čárkovanou vertikálou označenou A_{max}^d . Regresní rovnice $Q_s^d = a_{qs} + b_{qs} \cdot A^d$ a $Q_i^d = a_{qi} + b_{qi} \cdot A^d$ umožňují odhady spalného tepla a výhřevnosti pouze v rozpětí intervalu (A_{min}^d, A_{max}^d) . Pro odhad hodnot odpovídajících organické hmotě se musí rovnice upravit na tvar:

$$Q_{so}^d = a_{qs} + b_{qs} \cdot A^d - [(Q_{sM}^d : A_{max}^d) \cdot A^d] \text{ a poté } Q_{so}^o = (Q_{so}^d \cdot 100) : (100 - A^d \cdot f_A)$$

Analogicky se rovnice upraví pro odhad výhřevnosti organické hmoty. Pokud však jsou k dispozici konstanty posunutí a_{qs} , a_{qi} , je podstatně jednodušší způsob součinů:

$$Q_{so}^o = a_{qs} \cdot f_A \quad Q_{io}^o = a_{qi} \cdot f_A$$

Jak bylo již ukázáno, lze tento jednoduchý postup uplatnit i u jiných znaků jakosti uhlí.

Příklad odhadů složení organické hmoty a popelů z popelovin výběru n60

Obsah kyslíku organické hmoty byl odhadnut dopočtem do 100% po odhadu ostatních prvků pomocí součinnů $Y_o^o = a_Y \cdot f_A$; $C_o^o = 73,27$; $H_o^o = 5,40$; $S_o^o = 0,58$; $N_o^o = 1,12$; $O_o^o = 19,63$.

Spalné teplo a výhřevnost organické hmoty opět součinem $Y_o^o = a_Y \cdot f_A$; $Q_{so}^o = 30,57 \text{ MJ/kg}$; $Q_{io}^o = 29,36 \text{ MJ/kg}$.

Obsah oxidů v popelu vázaných popelovin: $SiO_2^A = 0$; $TiO_2^A = 0,69$; $Fe_2O_3^A = 25,27$; $Al_2O_3^A = 8,36$; $CaO^A = 29,58$; $MgO^A = 4,96$; $SO_3^A = 23,38$; $P_2O_5^A = 0,06$; $MnO^A = 0,08$; $Na_2O^A = 2,25$; $K_2O^A = 1,12$; $+(CO_2^A + \text{halogenáty}^A \text{ a další})$.

Obsahy oxidů v popelu volných popelovin: $SiO_2^A = 56,02$; $TiO_2^A = 2,87$; $Fe_2O_3^A = 3,55$; $Al_2O_3^A = 34,00$; $MgO^A = 0,59$; $MnO^A = 0,03$; $CaO^A = 1,36$; $Na_2O^A = 0,67$; $K_2O^A = 0,84$; $P_2O_5^A = 0,06$.

Obsahy oxidů v popelu vázaných a volných popelovin byly odhadnuty pomocí regresních rovnic $OX_i^A = a_{ox} + b_{ox} \cdot (1/A^d)$ a odhad každého oxidu vázaných popelovin byl proveden dosazením $(1:A_{min}^d)$, volných pak dosazením hodnot $(1:A_{max}^d)$.

3.1.5 Odběr vzorků

Slojové vzorky se obvykle odebírají na základě vizuálního makropetrografického rozlišení podle ČSN 44 1302 [17] z vrtných jader (segmentů nebo drti) nebo záseků. Vzorky lze získat též z těžného uhlí (kusy rozměrů cca 80 - 120mm), které se rozdělí podle přibližně odhadnutého obsahu popela pomocí zdánlivé hustoty stanovené hydrostatickým vážením (na vzduchu a ve vodě), případně radiometrickým měřením.

Je známo, že mezi reciprokou hodnotou zdánlivé hustoty (zdánlivým měrným objemem) a obsahem popela (v původním stavu) existuje lineární závislost

$$(1/d_a^r) = a + b \cdot A^r \quad (3.22)$$

Obsahy popela lze odhadnout podle vzorce

$$A^r \approx \left[\left(\frac{1}{a^r} \right) - a \right] : b \quad (3.22a)$$

kde např. pro černé uhlí OKD $a \approx 0,79$; $b \approx -0,005$

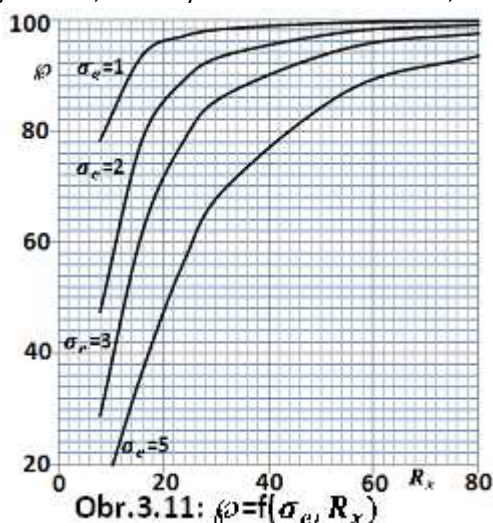
pro hnědé uhlí SHD $a \approx 0,86$; $b \approx -0,005$

pro hnědé uhlí HDB-lom Sokolov $a \approx 0,92$; $b \approx -0,0055$

Konstanty regresních rovnic pro jednotlivé lokality lze přibližně odhadnout pomocí dostatečně rozsáhlého výběru vzorků prostřednictvím rovnic (3.22) po vyloučení odlehlých hodnot podle čl. 2.9.1.7. Podmínkou je dostatečně rozsáhlé variační rozpětí obsahů popela, např. $R_x \approx 70$. Zdánlivou hustotu lze stanovit též podle ČSN 44 1321 [18].

3.1.6 Požadované vlastnosti výběru vzorků uhlí

Přesnost hodnot vypočítávaných „odhadů“ parametrů regresních rovnic je závislá na řadě vlivů a jejich měnlivosti. Postup hodnocení analýz uhlí vyžaduje již při výběrech souborů vzorků uhlí zajištění určitých specifických podmínek, které umožní získávání výsledků hodnocení s vhodnými a přijatelnými vlastnostmi. U vrtných a zásekových vzorků je to co nejostřejší rozdělení vrstev rozdílného obsahu popela. Dále je to rozsah souboru odebraných vzorků, který má být minimálně o cca 20% rozsáhlejší, než je rozsah n_{min} (viz tab. 2.2, str. 14), aby byla rezerva pro vylučování vzorků s odlehlými hodnotami. Vyloučením odlehlých dat se často podstatně zmenší reziduální směrodatná odchylka, což vede k žádoucímu zvětšení koeficientu determinace $\varphi = 100r^2(\%)$, viz čl. 2.9.1.1. Dále je to variační rozpětí nezávisle proměnné R_x (A^d , resp. $1:A^d$), které má být co největší, jehož zvětšení opět vede k žádoucímu zvětšení koeficientu determinace φ . Přibližný vztah mezi koeficientem determinace φ , reziduální směrodatnou odchylkou s_e a variačním rozpětím nezávisle proměnné R_x je patrný z obr. 3.11, kde na fiktivních čtyřech souborech o rozsazích $n=34$, reziduálních směrodatných odchylkách základního souboru $\sigma_e=1,2,3,5$ a variačních rozpětích $R_x=8$ až 80, je patrný vliv hodnot σ_e, R_x různé velikosti na hodnotu φ . U použitých fiktivních souborů jsou hodnoty x rovnoměrně rozděleny v celém variačním rozpětí, na rozdíl od náhodného rozdělení x v reálných souborech. U výběrů zbavených jednotek, zatížených nadlimitními residui, nemá být podkročen rozsah n_{min} . V opačném případě



Obr. 3.11: $\varphi = f(\sigma_e, R_x)$

výběry nutno doplnit nejméně na hodnoty n_{min} , přičemž k doplňování rozsahů výběrů lze vždy použít pouze hodnoty dat nezatížených nadlimitními reziduálními odchylkami. Vzorky mají být analyzovány co nejdříve po odběru, protože podléhají oxidaci a dalším změnám. Zacházení se vzorky viz ČSN 44 1304 [19].

3.1.7 Využití výsledků hodnocení.

Výsledné odhady parametrů lin. regresních rovnic je třeba hodnotit z hlediska zadaného účelu a předem určit, bude-li použit soubor po vyloučení odlehlých hodnot, nebo po doplnění rozsahu přijatelnými daty, či zda je vhodnější soubor neupravený, zahrnující všechny původní vzorky z rozdělených vrstev vrtu nebo záseku,

nebo použít vážených odhadů v případech velikých rozdílů mezi jednotlivými mocnostmi dělených vrstev. Volba takového rozhodnutí není pouze věcí matematickou, ale vyžaduje posouzení jak z hlediska chemického a fyzikálního, tak také geologického a technologického. Rozhodujícím momentem zde je samozřejmě účel, kterému má toto hodnocení sloužit a požadovaná přesnost odhadu. Tato upozornění mají především zdůraznit, že výběrovými soubory nelze nakládat zcela libovolně a bez rozmyslu, podobně jako nelze ekvivalentně nahradit experimentální hodnotu hodnotou regresní. Vždy je třeba zvažovat reprezentativnost výběrů a výsledků odhadů, které se blíží průměrným hodnotám podle zvoleného obsahu popela.

3.1.8 Postup při výpočtech

Pro časovou náročnost výpočtů, zejména při velkých rozsazích výběrů, je vhodné provádět výpočetní operace pomocí programu a samočinného počítače s programovatelnou tiskárnou. Je výhodné pro každý hodnocený výběr po vytištění odhadnutých parametrů regresní závislosti (např. tab. 2.5) a charakteristik (čl. 3.1.3) vytisknout i údaje pro jednotlivé závislosti, jak uvádí tabulka 3.2 s naznačenými fiktivními hodnotami.

Tabulka 3.2

Organizace:			Katastr:			Lokalita:		
Způsob vzorkování:			Počet vzorků:			Regresní funkce:		
Číslo vzorku	Hloubka v m	Mocnost v m	Argument x	Závislá y=Y+e	Funkce Y=f(x)	e (y-Y)	$ e \geq 2 \cdot s_e$ (y-Y)	$ e \geq 2,6 \cdot s_e$ (y-Y)
91463	145,80-146,00	0,20	0,01328	64,43	54,37	10,06	10,06	
91464	146,00-147,00	1,00	0,02145	59,95	53,44	6,51		
91465	147,00-147,01	0,01	0,13477	28,37	40,61	-12,24	-12,24	
91466	147,01-147,06	0,05	0,04114	31,15	51,21	-20,06	-20,06	-20,06
.....								

Ucelená tabulka poskytne přehled o stavu (hodnotě, rozložení, variabilitě) hodnocené složky nebo vlastnosti uhlí ve sloji, případně souslojí a možnostech případné selektivní těžby atd.

Literatura 3. kapitoly

- [1] KRYM, V.S.: Chimija tverdogo topliva, Charkov, ONTI-NKTI, 1934
- [2] KING, J.G. – MARIES, M.B. - CROSSLEY, H.E.: J.Soc.Chem.Ind., 55, 1936, 2771.
- [3] HUBÁČEK, J.- LUSTIGOVÁ, M.: Výpočet a stanovení uhelných popelovin, Zpráva ÚVP
Práce ÚVP, sv.9, Praha 1964.
- [4] ŠINDELÁŘ, J.: Paliva 37, 11, 381, 1957.
- [5] KOLEKTIV PRACOVNÍKŮ: Tabulky a diagramy z oboru paliv.Díl1, SNTL Praha 1956,232.
- [6] KONTA, J.: Jílové minerály Československa. NČSAV, Praha 1957.
- [7] ČSN 44 1364 Tuhá paliva – Výpočet obsahu popelovin, hydrátové vody a hodnot znaků
jakosti v organické hmotě – prosinec 1986.
- [8] BROWN, H. – DURIE, R.A. – SHAFER, H.W.S.: Fuel, 38, 1959, 295.
- [9] RADMACHER, W. – MOHRHAUER, P.: Brennstoff Chem., 36, 1955, 236.
- [10] GLEIT, Ch.E. - HOLLAND, W.D.: Anal. Chem., 34, 1962, 1454.
- [11] MACKOWSKA, M.Th.: Mitt. d. VGB, 38, 1955, 746.
- [12] STANSFIELD, E. – SUTHERLAND, J.W.: Coal Division. New York 1930.
- [13] ROMOVÁČKOVÁ, H. – KESSLER, F.: Zpráva ÚVVP, úkol 01-063-54-55/216, Praha 1954.
- [14] FIELDNER, A.C. – SELWIG, W.A.: Trans. AIME, 88, 1930, s.597.
- [15] ŠPETL, F.: Chem. Listy, 28, 1934, s. 118, 135, 151.
- [16] ČSN 44 1310 Tuhá paliva - Označování analytických ukazatelů a vzorce přepočtů na různé
stavy paliva – leden 2001.
- [17] ČSN 44 1302 Tuhá paliva – Směrnice pro vzorkování uhelných slojí – září 2001.
- [18] ČSN 44 1321 Tuhá paliva – Stanovení zdánlivé hustoty uhlí – březen 1985.
- [19] ČSN 44 1304 Tuhá paliva – Metody odběru a úpravy vzorků pro laboratorní zkoušení. Duben
1989.
- [20] KUBANT, J.: Studie možnosti vyjádření závislosti složení popela jako funkce obsahu popela v
uhlí. Zpráva ÚVP, úkol 71 392/R, Praha 1972.
- [21] HUBÁČEK, J. a kol.: Chemie uhlí., SNTL Praha/Bratislava, 1962.
- [22] HAVLENA, V.: Geologie uhelných ložisek 1., NČSAV, Praha 1963.
- [23] RUSČEV, D.D.: Chimija tverdogo topiva. Izd. Chimija, Leningrad 1976.
- [24] BOUŠKA, V.: Geochemie uhlí. Academia, Praha 1977.
- [25] GUMZ, W. – KIRSCH, H. – MACKOWSKA, M.Th.: Schlackenkunde., Springer Verl.
Berlin/Göttingen/Heidelberg 1959.
- [26] WALL, T.F. – LOWE, A. – WIBBERLEY, L.J. - STEWART, I.Mc.C.: Mineral matter in coal and the
thermal performance of large boilers. SPECS Vol.5, No.1., A 1979, 29,s.
Pergamon Press Ltd., GB.
- [27] KUBANT, J.: Stanovení skutečného obsahu hořlaviny ve vysokopopelnatých palivech. Úkol
ÚVP 61 208/I, Praha 1969.
- [28] ČSN 44 1322 Tuhá paliva – Stanovení skutečné hustoty tuhých paliv 12.1967.