

METODY RYCHLÉHO VYHODNOCOVÁNÍ EXPERIMENTÁLNÍCH DAT

Ing. Oldřich Kropáč, CSc

Nedílnou součástí každé experimentální činnosti je vyhodnocení experimentálně získaných údajů. Racionální přístup k řešení současných problémů vědy a techniky vyžaduje, aby trvalé úsilí o zlepšování metodiky každého experimentálního oboru bylo doprovázeno zdokonalováním a prohlubováním metodiky zpracování získaných dat. Teoretickou základnu k tomu poskytuje teorie pravděpodobnosti a matematická statistika /1/. Mezi praktickými experimentátory mnohdy ještě panuje názor, že přesné vyhodnocovací postupy jsou příliš složité pro každodenní rutinní aplikace. Účelem tohoto krátkého pojednání je ukázat, že ze statistické teorie lze při použití některých dalších jednoduchých analytických a grafických prostředků odvodit vyhodnocovací postupy, které jsou jednoduché, rychlé a zvládnutelné pomůckami dostupnými každému jednotlivému experimentátorovi.

1. ZÁKLADNÍ POJMY A POUŽITÉ METODICKÉ PŘÍSTUPY

Pro rychlé vyhodnocení experimentálních dat, s cílem získat odhady distribučních funkcí, budeme využívat účelnou kombinaci tří metodických přístupů, které budeme nyní stručně charakterizovat.

Především je to úprava experimentálních dat vhodnými transformacemi. U deterministických veličin často napomáhají transformace souřadných soustav ke zjednodušení analytických vztahů, např. k linearizaci funkčních závislostí.

Po transformaci $\varphi(X) = Z$ náhodné veličiny X s hustotou pravděpodobnosti $f_X(x)$ má výsledná náhodná veličina Z hustotu pravděpodobnosti

$$f_Z(z) = f_X[\varphi(z)] \cdot d\varphi(z)/dz, \quad (1.3)$$

kde $\varphi(Z) = X$ je funkce inverzní k $Z = \varphi(X)$.

Nejčastěji využíváme transformace experimentálních dat k přibližné normalizaci dat (resp. zmenšení součinitele šikmosti), zejména je-li náhodná veličina definována na polopřímce (podrobnosti viz např. /4/ ÷ /8/) a v regresních úlohách, chceme-li odstranit závislost rozptylu na očekávané hodnotě (je-li $\sigma \approx E^m$, je vhodnou transformací $z = x^{1-m}$; je-li $m = 1$, pak $z = \log x$).

Johnson /9/ ukázal, že na normální rozdělení lze převést spojitě distribuční funkce se širokým rozmezím součinitelů šikmosti a špičatosti, a to pomocí pouze tří transformačních vztahů, a to: $z = \log x$ pro rozdělení lognormální, $z = \log(x/(1-x))$ pro rozdělení oboustranně omezené, tzv. typ S_B , $z = \log(x + \sqrt{x^2 + 1})$ pro rozdělení oboustranně neomezené, typ S_U . Johnsonovy transformace jsou vhodné zejména při analýze sdružených negaussovských rozdělení /10/.

Pro rychlé vyhodnocení různých statistik z relativně malých souborů experimentálních dat je výhodné uspořádat daný náhodný výběr $\{x_i\}$, $i = 1, 2, \dots, n$ do vzestupné posloupnosti $x_{(1)} \leq x_{(2)} \leq x_{(3)} \dots x_{(i)} \dots \leq x_{(n)}$. Konkrétní návody včetně tabulek statistických veličin, které se přitom používají, lze nalézt v citovaných publikacích /11/, /12/.

Jako třetí metodický přístup, který je velmi účinný a v různých aplikačních oborech, zejména technických, velmi oblíbený, je používání grafických metod, a to nejen pro úsporné a přehledné zobrazení výsledků, ale též ke grafickému řešení jednodušších úloh. Pokud jde o úlohy z oblasti matematické statistiky, lze zejména rychle a jednoduše testovat některé hypotézy, jestliže se převedou na geometrickou úlohu posouzení souladu souboru experimentálně získaných bodů s vhodně situovanou přímkou.

2. VYHODNOCENÍ DISTRIBUTUČNÍ FUNKCE A JEJICH PARAMETRŮ

Současný rozvoj metodiky vyhodnocování experimentálních údajů většiny experimentálních oborů dospěl do stadia, kdy již nestačí odhady očekávané hodnoty, popř. rozptylu, nýbrž je nutno najít typ distribuční funkce, s kterou jsou experimentální data v souladu a najít odhady parametrů této distribuce. Ukážeme si, že tuto úlohu lze snadno řešit, jestliže účelně zkombinujeme metodické přístupy stručně charakterizované v kap.1. Pro přehled celý postup rozčleníme na dílčí kroky.

2A. Posouzení shody empirické distribuční funkce s předpokládaným analytickým tvarem

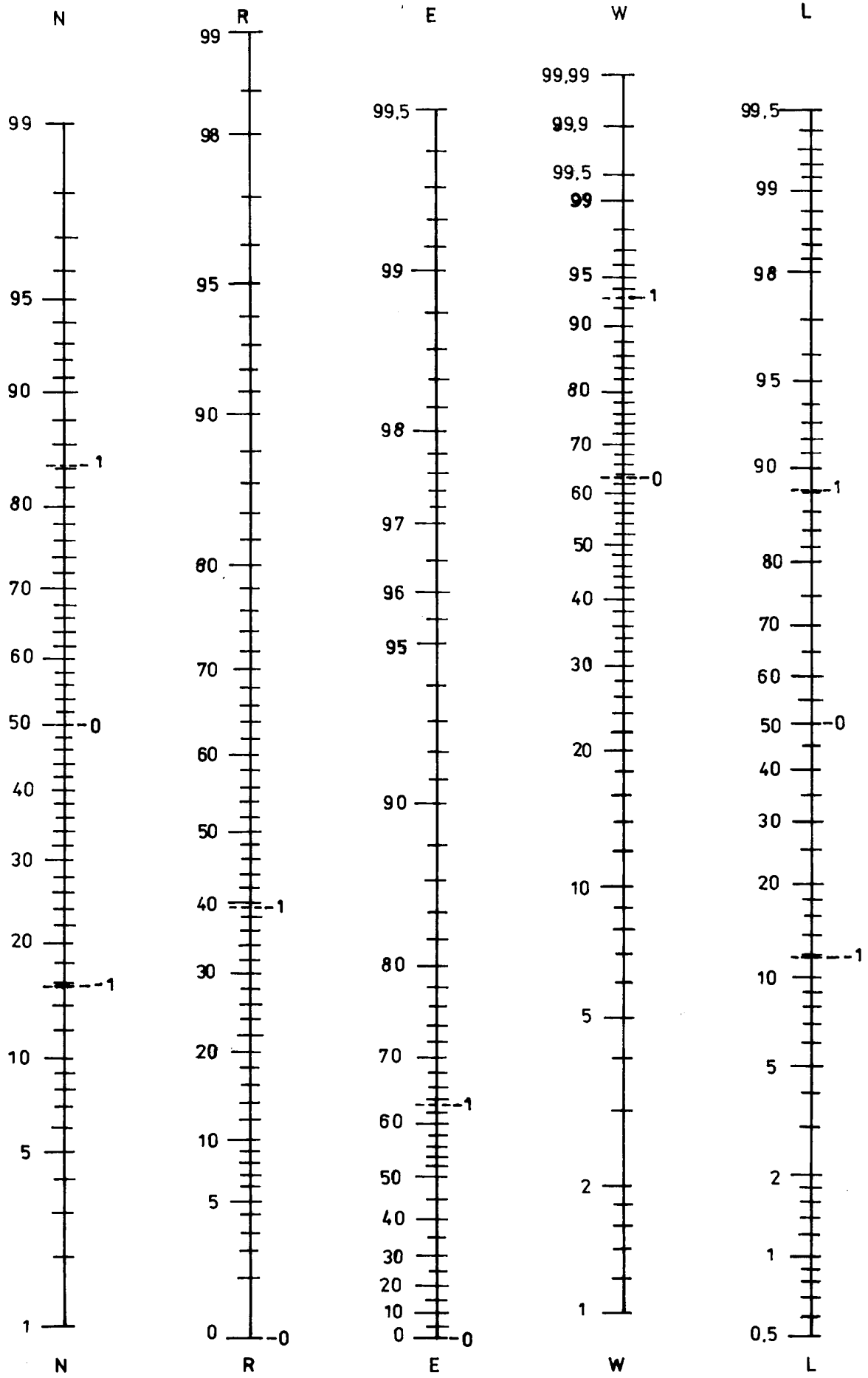
Z praktických zkušeností vyplývá, že vizuálně v grafu se shoda dvou závislostí vůči sobě nejlépe posoudí, jestliže posuzujeme shodu s přímkou. To lze v našem případě použít tak, že empirickou

distribuční funkci $F(X)$ vyneseme do tzv. pravděpodobnostního papíru. To je graf, jehož stupnice jsou upraveny tak, aby distribuční funkce (které pravděpodobnostní papír přísluší) se v něm zobrazila jako přímka. Vyneseme-li do takového papíru empirickou distribuční funkci, zjistíme ihned přiložením pravítka, zda body podle (2.1) resp. podle (2.2) jsou rozloženy kolem přímky, nebo zda se od ní výrazně odchyľují.

Pravděpodobnostní papíry lze sestrojít pro rozdělení pravděpodobnosti, jejichž distribuční funkce lze transformacemi převést na lineární vztah mezi dvěma parametry, a to buďto mezi parametrem polohy a parametrem měřítka, popř. mezi parametrem měřítka a parametrem tvaru (přičemž parametr polohy musí být znám). Teorie a praxe pravděpodobnostních papírů je podrobně popsána v člancích /14/ ÷ /26/ (uspořádaných chronologicky), v češtině je souborně zpracována v /27/ ÷ /29/. Na obr.1 je uvedeno pět stupnic pravděpodobnosti pro často používaná rozdělení. Kombinací stupnic pro pravděpodobnosti $y(P)$, kde $P \equiv F(X)$ a stupnic experimentálních dat X můžeme vytvořit soubor pravděpodobnostních papírů pro rozdělení často používaná v experimentální praxi. Nejdůležitější z nich jsou shrnuta v tabulce 1. Pro rozdělení Gumbelovo (resp. Fréchetovo) se dá použít obrácená stupnice rozdělení Weibullova. Jak z tabulky 1 vyplývá, platí totiž $y(P)_{\text{Gumbel}} = -y(1-P)_{\text{Weibull}}$.

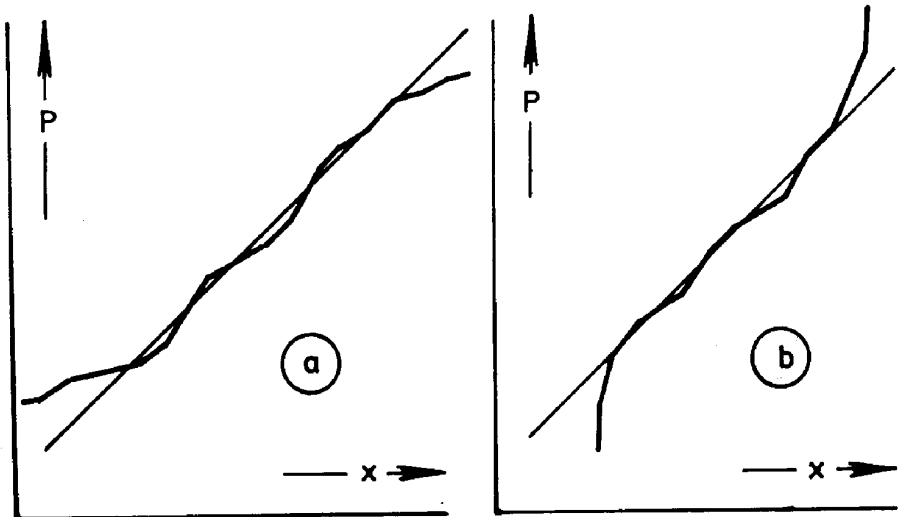
Tabulka 1. Stupnice pravděpodobnostních papírů vybraných rozdělení

Stupnice pravděpodobností $y(P)$	Stupnice experimentálních dat		
	x	$\ln x$ ($\log x$)	jiné
P	rovnoměrné		
$u_P =$ kvantily normálního rozdělení	normální	lognormální	Johnsonovy transformace: S_U S_B
$-\ln(1-P)$	exponenciální (Laplaceovo)	Paretovo	x^2 : Rayleighovo
$\sqrt{-2\ln(1-P)}$	Rayleighovo		
$\ln(-\ln(1-P))$		Weibullovo	
$-\ln(-\ln P)$	Gumbelovo	Fréchetovo	
$\operatorname{arctgh}(2P-1)$	logistické		
$\operatorname{tg}[\pi(P-0,5)]$	Cauchyovo		



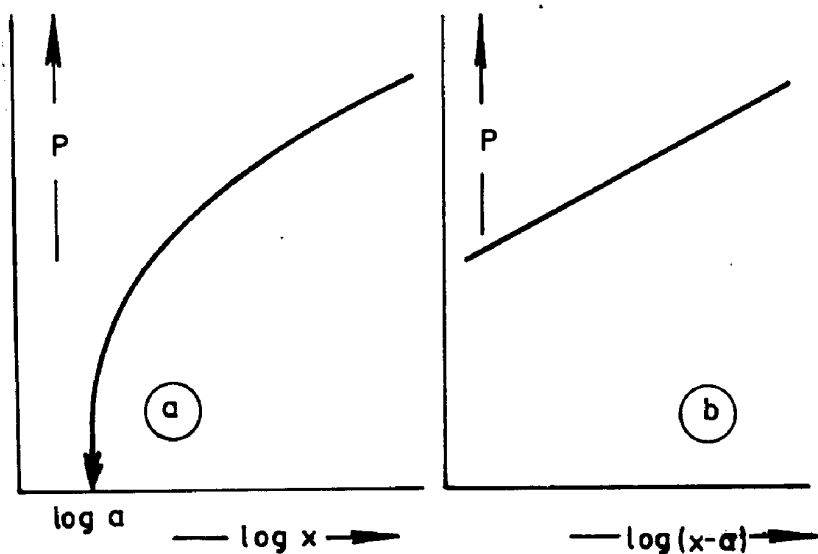
Obr.1.

Stupnice pravděpodobností v % pro pravděpodobnostní papíry rozdělení normálního (N), Rayleighova (R), exponenciálního (E), Weibullova (W) a logistického (L).



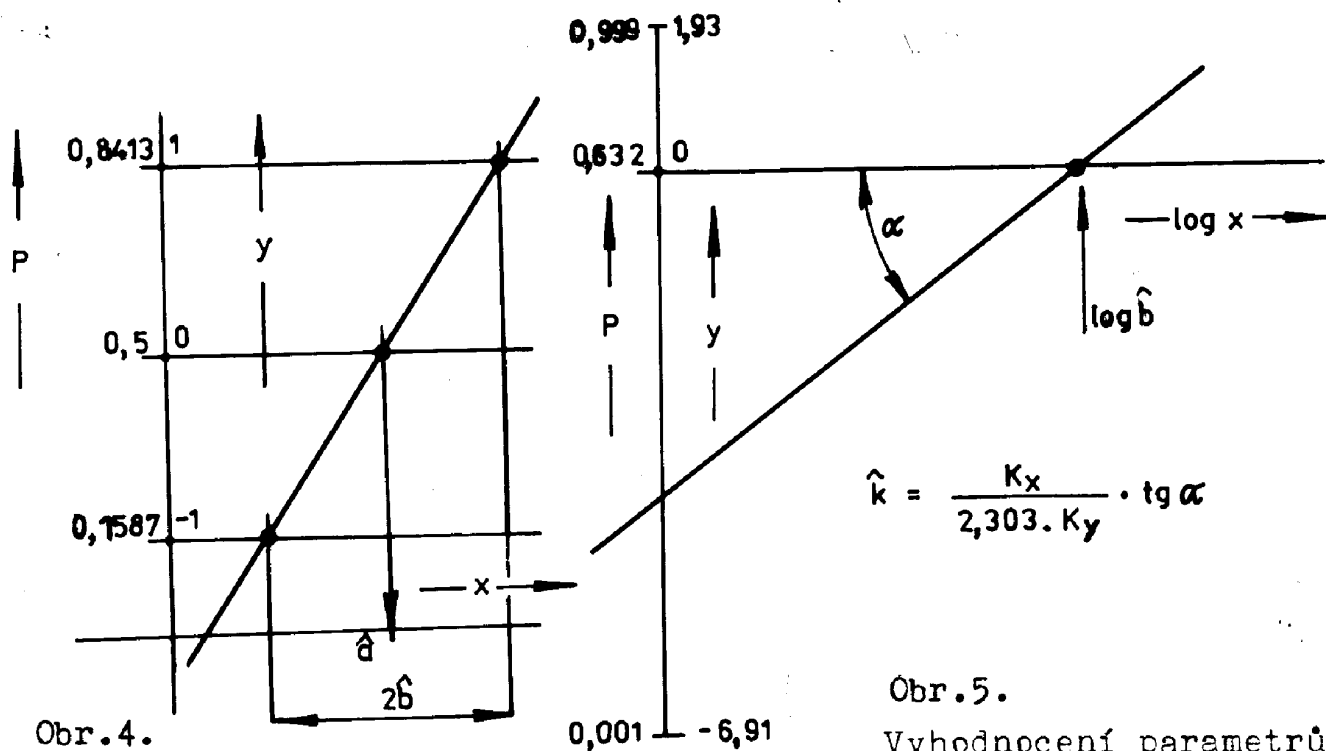
Obr.2.

K výkladu příčin odchylek okrajů empirické distribuční funkce od přímky



Obr.3.

Vliv přehové hodnoty na distribuční funkci v pravděpodobnostním papíře s logaritmickou stupnicí náhodné proměnné



Obr.4.

Vyhodnocení parametrů normálního rozdělení z jeho pravděpodobnostního papíru

$$\hat{k} = \frac{K_x}{2,303 \cdot K_y} \cdot \operatorname{tg} \alpha$$

Obr.5.

Vyhodnocení parametrů Weibullova rozdělení z jeho pravděpodobnostního papíru

Jestliže se empirická distribuční funkce výrazně odchyluje od přímky celkovým charakterem svého průběhu, je nutno změnit předpoklad o typu rozdělení a použít jiný pravděpodobnostní papír. Někdy je průběh empirické distribuční funkce přibližně přímkový v téměř celém oboru pravděpodobnosti, s výjimkou některého (popř. obou) konců rozdělení. Jsou-li odchylky konců podle obr.2a (zejména jde-li o pouze jednu nebo dvě krajní hodnoty), jde zpravidla o odlehlé hodnoty, které se do prokládání přímky neuvažují. Mají-li odchylky od přímky charakter podle obr.2b, jde o rozdělení useknuté. Objeví-li se průběh podle obr.3a na grafu s logaritmickou stupnicí náhodné proměnné, znamená to, že existuje nenulový parametr polohy a (který zdemá význam "prahové" hodnoty), k níž limituje levý okraj empirické distribuční funkce, takže parametr polohy se dá z grafu odečíst. Po transformaci $X' = X - a$ leží již obvykle empirická distribuční funkce $F(x')$ ve zvoleném papíře na přímce (obr.3b).

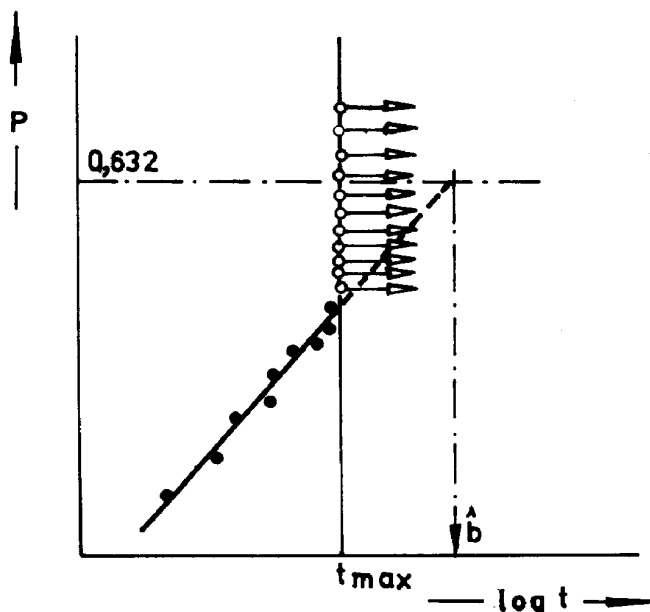
2.2. Odhad parametrů rozdělení

Jestliže jsme zjistili dobrý soulad empirické distribuční funkce s přímkou, můžeme z ní graficky odhadnout parametry dotyčného rozdělení. Grafické vyhodnocení parametrů normálního rozdělení je ukázáno na obr.4, vyhodnocení parametrů Weibullova rozdělení je ukázáno na obr.5. Pro vyhodnocení parametrů přímky vyrovnávající průběh empirické distribuční funkce se používají hodnoty P , pro něž $y(P)$ nabývají hodnoty 0 nebo 1, resp. -1 (viz obr. 4,5). Tyto hodnoty jsou ve stupnicích na obr.1 vyznačeny slabě čárkovane a jejich hodnoty jsou připsány po pravé straně stupnic. Pro rychlé vyhodnocení parametrů z pravděpodobnostních papírů slouží i jednoduché nomografické pomůcky, které jsou často součástí předtištěných rastrů, např. pro vyhodnocení parametrů Weibullova rozdělení /20/.

Metoda odhadování empirických distribučních funkcí založená na uspořádání experimentálních dat umožňuje zpracovávat i velmi malé výběry (v literatuře se uvádí od $n = 6$). V těchto případech však při použití vztahu (1.4) pro pravděpodobnost $P_{(i)}$ přiřazenou hodnotě $x_{(i)}$ nejsou získané odhady parametrů zcela nestranné a vykazují i sníženou efektivnost. K zlepšení této situace se pro rozdělení normální a Gumbelovo doporučuje vztah $P_{(i)} = (i-0,375)/(n+0,25)$ /18/, popř. se studují i další možnosti co nejlépe uspořádaného rozložení úrovní $P_{(i)}$.

Možnost získání odhadů distribučních funkcí i z malých souborů dat může znamenat při náročných experimentech (např. drahé zkušební vzorky) značný ekonomický přínos.

U zkoušek na únavu nebo životnost přistupují navíc nemalé náklady za dlouhodobou realizaci zkoušek (týdny až měsíce) a kromě toho se takto dlouho



Obr.6.

Schéma vyhodnocení cenzurovaného rozdělení (Weibullova)

souboru výsledků do pravděpodobnostního papíru dostaneme dvě skupiny bodů, viz obr.6. Body zobrazující porušené vzorky budou v levé spodní části diagramu ležet přibližně na šikmé přímce (černé body), body zobrazující nepoškozené vzorky budou mít své pořadové pravděpodobnosti, avšak pro nedostatek další informace je zakreslíme na úroveň t_{max} (prázdné kroužky) a neporušení vyznačíme šipkami vpravo od t_{max} . Pro vyhodnocení použijeme pouze body příslušející porušeným vzorkům a pro získání odhadu parametrů použijeme v případě potřeby lineární extrapolaci (viz obr.6). Jako extrémní případ cenzurování lze uvést postup použitý v oboru kuličkových ložisek, kde se do zkoušek dává současně 30 vzorků a zkoušení se ukončí, jakmile dojde k poruše první ^{ch} tří vzorků /21/. Podrobnosti složitějších případů cenzurování jsou uvedeny v článku /24/.

Je samozřejmé, že odhadování parametrů, které jsme popsali graficky, lze realizovat pomocí přímkové regresní analýzy na počítači. Podrobnosti tohoto postupu jsou uvedeny např. v /13/, /25/, /27/ i jinde. Připomeňme pouze, že za nezávisle proměnnou ("přesnou") se považuje $P_{(i)}$, za závisle proměnnou zatíženou náhodnými chybami se považuje $x_{(i)}$. Při nelineární transformaci proměnné $x_{(i)}$ je dále nutno respektovat váhy vyplývající z použité transformace.

trvajícím experimentem snižuje aktuálnost nakonec získaného výsledku. Proto se realizují tzv. cenzurované zkoušky. Daný soubor vzorků zkoušíme pouze určitou dobu t_{max} . Během této doby se část vzorků porušila, přičemž jsme zaznamenali doby t_i , při kterých se to stalo. O zbývajících vzorcích máme pouze kvalitativní informaci, že jejich životnost je větší než t_{max} .

Při vynesení takového sou-

3. REGRESNÍ ANALÝZA

Pro experimentální výzkum jsou zejména důležité úlohy, v nichž vystupuje větší počet proměnných, ať již charakteru náhodného nebo zčásti deterministického. I když se při vyhodnocování takových úloh zpravidla obrátíme na samočinný počítač, je řada možností, jak některé základní úlohy řešit jen s elementárními pomůckami a přitom rychle a v principu objektivně, i když jen s určitou přibližností. Pro značnou rozsáhlost problematiky se omezíme jen na několik vybraných postupů.

3.1. Sdružené náhodné veličiny

Budeme uvažovat spojitou náhodnou veličinu o dvou složkách (X, Y) se sdruženou hustotou pravděpodobnosti $f(x, y)$, kterou vyjádříme jako součin marginální hustoty $f_x(x)$ a podmíněné hustoty $f_{y|x}(y|x)$, tedy

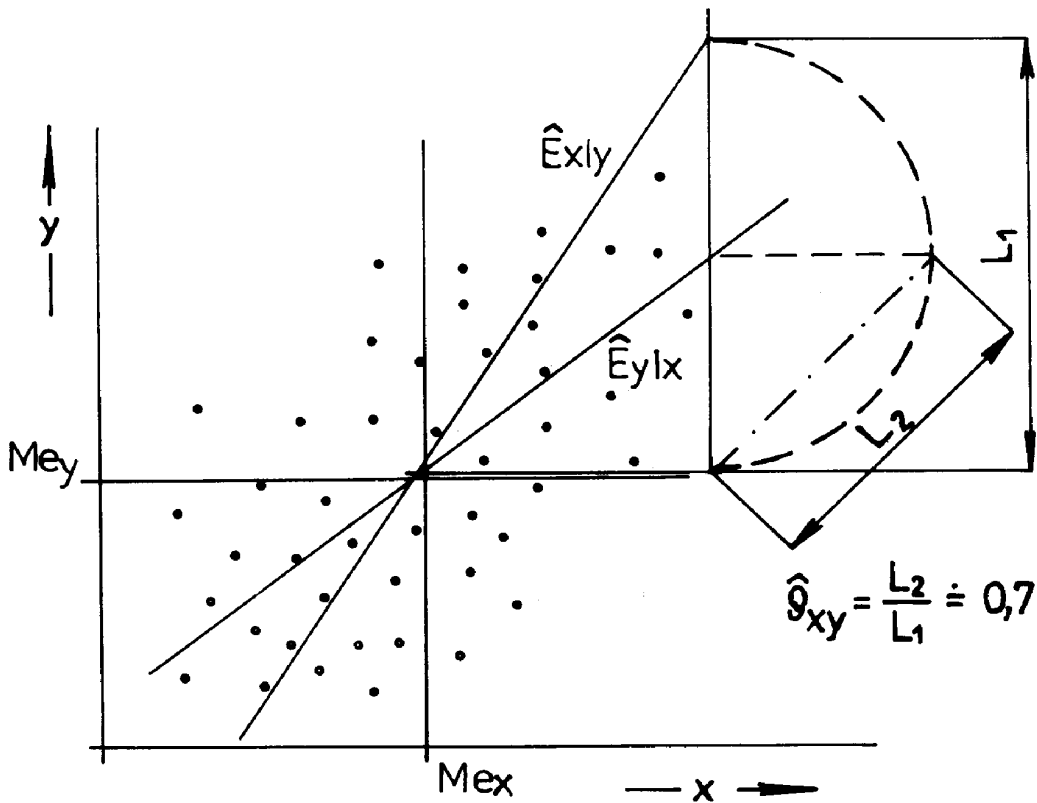
$$f(x, y) = f_x(x) \cdot f_{y|x}(y|x). \quad (3.1)$$

Výsledky experimentu na dvousložkové náhodné veličině (X, Y) , které označíme $\{x, y\}_i$, $i=1, 2, \dots, n$ pro názornost obvykle zakreslíme do tzv. bodového diagramu. Lze-li předpokládat, že $f_x(x)$ a $f_y(y)$ jsou symetrické a korelační vazba je lineární, nabízí se grafické vyhodnocení základních charakteristik souboru $\{x, y\}$ pomocí mediánových přímk. Přitom postupujeme takto:

Bodový graf (obr.7) rozdělíme mediánovou přímkou $x = Me_x$ na dvě části tak, aby v každé z nich byl stejný počet bodů. Pak nastavíme pravítko tak, aby jím vyznačená přímka rozdělovala levou polovinu bodů na dva stejně velké soubory a současně i pravou polovinu bodů na dvě stejně velké skupiny. Takto získáme mediánovou přímkou $Me_{y|x}$, kterou považujeme za odhad regresní přímky $\hat{E}_{y|x}$. Podobným způsobem najdeme odhad regresní přímky $\hat{E}_{x|y}$. Vzhledem k tomu, že mezi směrnici regresních přímk $b_{x|y}$ a $b_{y|x}$ platí vztah

$$b_{x|y} \cdot b_{y|x} = \rho_{xy}^2, \quad (3.2)$$

můžeme z regresních přímk pomocí Euklidovy věty snadno vyhodnotit odhad součinitele korelace. Grafická konstrukce je zřejmá z obr.7.



Obr.7.

Grafická konstrukce mediánových přímek a odhadu součinitele korelace

Je-li bodový graf příliš obsáhlý, rozdělíme korelační pole podle obou proměnných na pásma obsahující postupně 1/4, 1/2 a 1/4 výsledků. Při prokládání mediánových přímek se postupuje tak, že na poloviny se dělí jen obě krajní pásma (po čtvrtinách výsledků), kdežto střední pásmo se neuvažuje. Další varianty grafického vyhodnocení korelačních vztahů lze nalézt v /30/.

Je třeba připomenout, že nejsou-li splněny předpoklady této metody (rozdělení $f_x(x)$, $f_y(y)$ jsou souměrná a korelační vazba je lineární), budou graficky vyhodnocené odhady více či méně odchýlné od odhadů získaných výpočtem.

Pro podrobnou specifikaci sdruženého rozdělení na základě souboru experimentálních dat $\{x,y\}_i$ můžeme použít upravenou verzi postupu popsaného v kap.2. Metodické podrobnosti jsou uvedeny v /28/, /29/.

3.2. Náhodná veličina závislá na deterministické proměnné

Ve fyzikální a inženýrské experimentální praxi je velmi častý případ, kdy mezi dvěma deterministickými veličinami, a to mezi nezávisle proměnnou t a závisle proměnnou x platí deterministický

vztah $x = g(t)$, avšak při experimentálním ověřování tohoto vzta-
hu je závisle proměnná zatížena významnou náhodnou chybou \check{X} , za-
tímco nezávisle proměnnou t lze podle konkrétních okolností zjiš-
ťovat přesně nebo (ve vztahu k rozmezí hodnot této veličiny) jen
s nepatrnými náhodnými chybami. Pro takový případ použijeme analy-
tický model měřené veličiny X ve tvaru

$$X = g(t) + \check{X}, \quad (3.3)$$

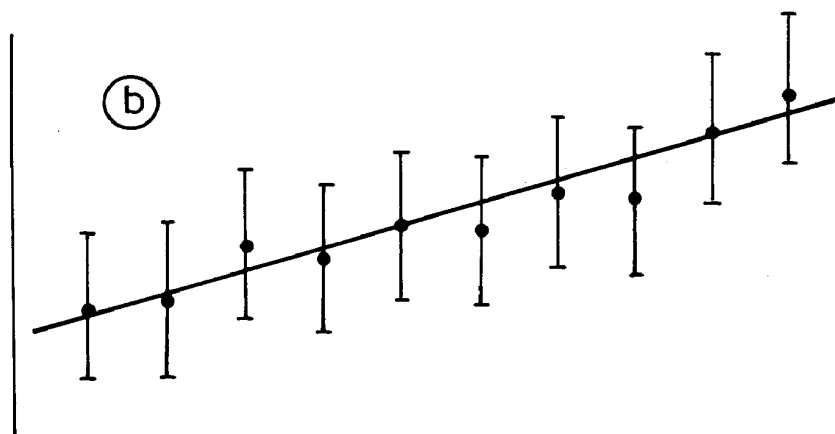
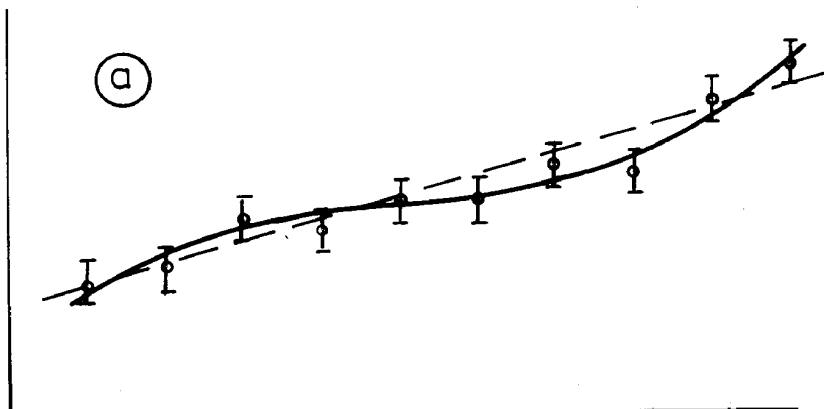
přičemž obvykle předpokládáme, že \check{X} je $N(0; \sigma_X)$. Pokud není \check{X} nor-
mální, použijeme vhodnou transformaci podle návodů zmíněných
v kap.1. Vztah (3.) lze ekvivalentně zapsat ve tvaru $X:N(g(t); \sigma_X)$.
Lze rovněž připustit závislost $\sigma_X(t)$. Pro nastavování hodnot ne-
závisle proměnné t se předpokládá takový krok $\Delta t = t_{i+1} - t_i$,
aby veličiny (X_{i+1}, X_i) byly nekorelované.

Experimentální data označíme $\{t_i, x_{ij}\}$, $i=1,2,\dots,n$; $j=1,2,\dots,n_i$,
kde i je index pro počet úrovní, na nichž nastavujeme nezávisle
proměnnou t ,

j je index pro počet opakovaných měření veličiny X na úrovni t_i .
Naším úkolem je jednak najít odhad funkčního vztahu $g(t)$, jednak
odhad rozptylu σ_X^2 resp. $\sigma_X^2(t)$. Tato úloha se řeší obvykle nume-
ricky s použitím metody "nejmenších čtverců". Znamená to, že poža-
dujeme, aby byl minimální součet čtverců odchylek naměřených hod-
not x_{ij} od aproximovaných $g^*(t_i)$, kde hvězdičkou u $g(t)$ jsme vy-
značili, že byly vhodně odhadnuty volné parametry tohoto funkční-
ho vztahu.

Metoda nejmenších čtverců je popsána v každé učebnici nebo
příručce aplikované statistiky, viz např. /1/. Nejprve však uká-
žeme několik grafických postupů, které při dodržování základních
předpokladů, za nichž byly odvozeny, dají dobré odhady.

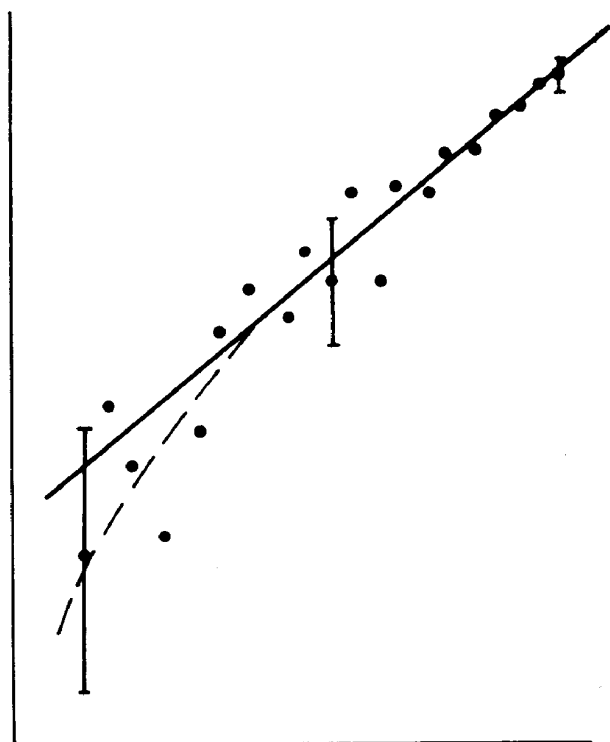
Velmi rychlé a oblíbené je grafické vyrovnávání experimentál-
ních výsledků "od ruky". Při posuzování předpokládaného průběhu
závislosti je nutné uvážit velikost experimentálního rozptylu od-
hadnutého z opakování na stejné úrovni nebo z přesnosti měřicí me-
tody či měřicího přístroje. Dvojnásobek odhadu směrodatné odchylky
jednotlivého měření vyneseme na obě strany od průměrné hodnoty na
dané úrovni. V případě vysoké přesnosti měření je nutno proložit
složitější křivku, která prochází v těsné blízkosti těchto bodů,
viz obr.8a. Je-li však rozptyl velký, má smysl použít pro vyrovná-
ní jen přímku (obr.8b). Při prokládání křivky se opět snažíme, aby
z experimentálních bodů ležela vždy zhruba polovina na každé stra-



Obr.8.
K otázce tvaru
vyrovnávající
funkce

ně křivky. Pokud jde o malý počet bodů (zejména je-li jen jedno opakování na každé úrovni) a konfidenční pásmo neposkytuje dosti výrazné rozlišovací kritérium, dbáme, aby se experimentální body

při projíždění křivkou střídaly po obou jejích stranách. Přesné statistické kritérium poskytuje tzv. test serií (viz např. /34/).



Obr.9.
Vliv proměnné přesnosti
měření na vyrovnání

Podobně jako u vyhodnocování distribučních funkcí, je i u úloh tohoto typu výhodné vynášet výsledky do grafů s funkčními stupnicemi (nejčastěji logaritmickými), kde se většina průběhů zobrazí přímkami. Přitom však musíme pamatovat, že v různých místech diagramu se ta či ona fyzikální jednotka zobrazí úsečkami různé délky. Pak pomůže vynést si kolem několika typických bodů (na okrajích a ve středu) úsečky

95% konfidenčního pásma podobně jako v obr.8 a vyrovnání udělat tak, aby vyrovnaný průběh procházel těmito úsečkami proporcionálně ve všech částech grafu. Z tohoto hlediska tedy v obr.9 postačí lineární vyrovnání a není třeba v levé části grafu uvažovat zakřivený průběh podle čárkované čáry.

Zajímavou grafickou metodu vyrovnávání experimentálních dat přímkou uvádí Dolby /31/. Vychází ze skutečnosti, že rovnici přímky $y = a + bx$ zobrazenou v souřadné soustavě (x,y) s parametry a, b , lze přepsat do tvaru $a = y + (-x).b$, kterou lze zobrazit v souřadné soustavě (a,b) s parametry $y, -x$. Souřadné soustavy (x,y) a (a,b) jsou navzájem duální a platí, že bodu v (x,y) odpovídá přímka v (a,b) a naopak. Odchytky ve svislém směru experimentálních bodů od vyrovnané přímky v (x,y) se zobrazí jako vzdálenosti (ve svislém směru) jednoho bodu od souboru přímek. Parametry přímky proložené experimentálními body tak, aby součet absolutních odchylek byl minimální, se proto lehce najdou v rovině (a,b) tímto postupem:

1. Experimentální body vyneseme do souřadných soustav (x,y) a (a,b) , viz obr.10. Všimněme si, že zvolíme-li stupnice pro y a a stejné s počátkem ve stejné výškové úrovni, promítneme souřadnice y_i z (x,y) rovnoběžně s osou x do pořadnic na ose a .

2. Najdeme minimální šířku obálky přímek v (a,b) ve směru osy a . Rozpůlením této šířky dostaneme bod P , jehož souřadnice a^* , b^* udávají pořadnici na ose y a směrnici vyrovnávající přímky p .

3. Promítnutím a^* z (a,b) rovnoběžně s osou b na osu y dostaneme bod, z něhož vychází vyrovnávající přímka se směrnici b^* .

Požadujeme-li přesné zpracování naměřených dat, je nutno se obrátit k numerickým metodám. I zde jsou však možnosti, jak rozsah numerických operací co nejvíce zredukovat.

Kriterium metody nejmenších čtverců se zapíše ve tvaru

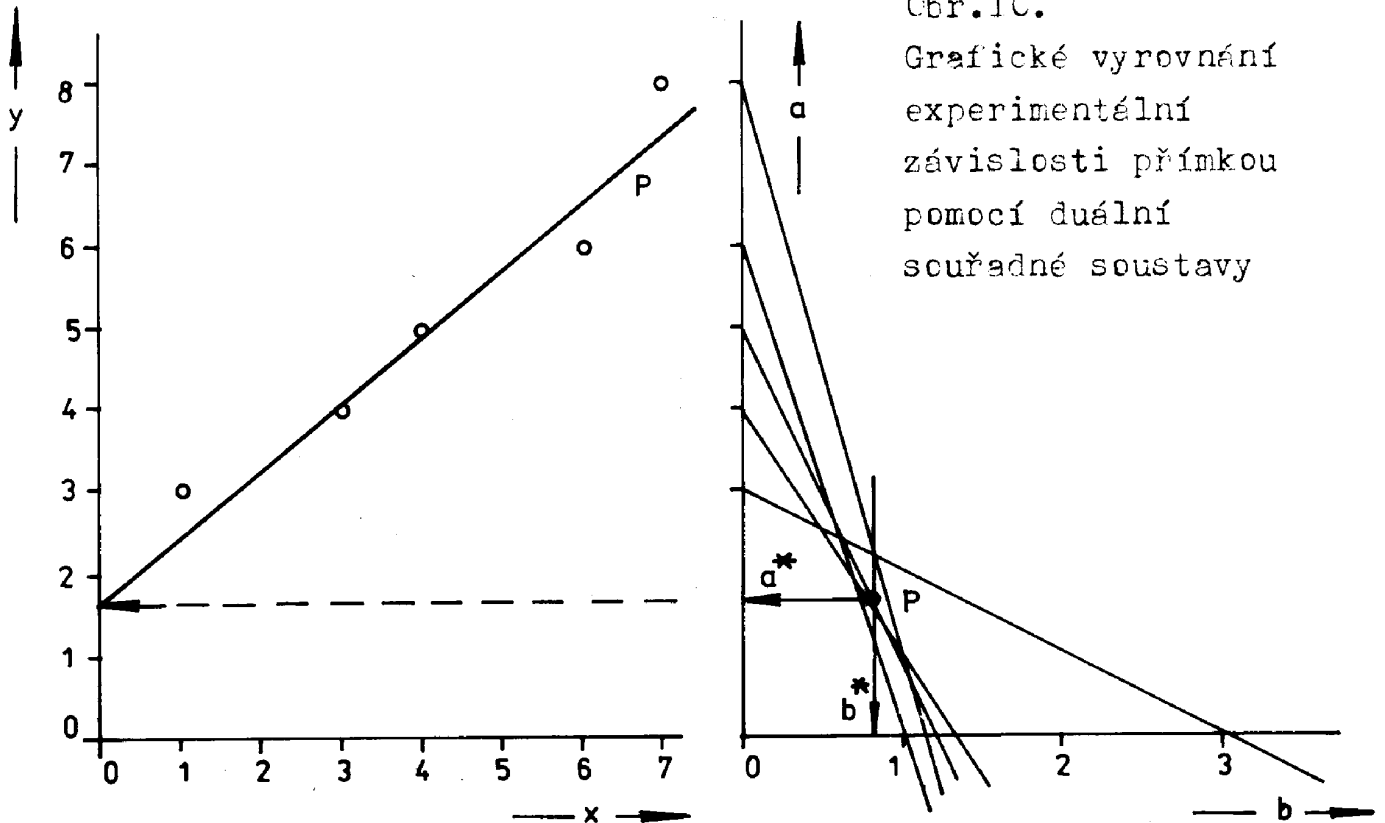
$$\sum_{i=1}^n \sum_{j=1}^{n_i} [x_{ij} - g(t_i)]^2 \longrightarrow \text{minimum.} \quad (3.4)$$

Funkční vztah $g(t)$ obsahuje kromě nezávisle proměnné t ještě zatím neurčené parametry $\alpha_0, \alpha_1, \dots, \alpha_q$, takže pro úplnost píšeme $g(t; \alpha_0, \alpha_1, \dots, \alpha_q)$. Podmínku (3.7) pak můžeme vyjádřit ve tvaru

$$\frac{\partial}{\partial \alpha_k} \sum \sum [x_{ij} - g(t_i; \alpha_0, \alpha_1, \dots, \alpha_k, \dots, \alpha_q)]^2 = 0; k = 0, 1, \dots, q \quad (3.5)$$

Obř.10.

Grafické vyrovnání
experimentální
závislosti přímkou
pomocí duální
souřadné soustavy



a po rozepsání

$$\sum \sum g(t; \alpha_0, \alpha_1, \dots) \cdot \frac{\partial g(t; \alpha_0, \alpha_1, \dots)}{\partial \alpha_k} = \sum \sum x_{ij} \cdot \frac{\partial g(t; \alpha_0, \alpha_1, \dots)}{\partial \alpha_k}, \quad (3.6) \quad k=0, 1, \dots, q$$

Pro konkrétní vyhodnocení je nutno zadat konkrétní tvar funkce $g(t; \alpha_0, \dots)$. Nejvýhodnější situace nastane, jestliže $g(t; \alpha_0, \alpha_1, \dots)$ je lineární ve vztahu k parametrům α_k , tj. můžeme ji vyjádřit ve tvaru

$$g(t; \alpha_0, \alpha_1, \dots, \alpha_q) = \sum_{k=0}^q \alpha_k \cdot g^{(k)}(t) \quad (3.7)$$

kde $g^{(k)}(t)$ jsou libovolné, navzájem lineárně nezávislé funkce, které neobsahují žádný volný (neznámý) parametr.

Budiž např. $g(t)$ mnohočlen q -tého stupně

$$g(t) = \alpha_0 + \alpha_1 \cdot t + \alpha_2 \cdot t^2 + \dots + \alpha_q \cdot t^q \quad (3.8)$$

a pro zjednodušení zápisu předpokládejme, že $n_i=1$ (na každé úrovni t_i je pouze jedno měření), takže v (3.6) zůstanou jen sumy $\sum_{i=1}^n$ které budeme psát zkráceně \sum . Pak soustavu (3.6) můžeme přepsat do soustavy tzv. normálních rovnic pro α_k , $k=0, 1, \dots, q$ (v maticovém zápise)

$$\begin{bmatrix} n & \sum t & \sum t^2 & \dots & \sum t^q \\ \sum t & \sum t^2 & \sum t^3 & & \vdots \\ \sum t^2 & \sum t^3 & \sum t^4 & & \vdots \\ \vdots & & & & \vdots \\ \sum t^q & \dots & \dots & \dots & \sum t^{2q} \end{bmatrix} \times \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_q \end{bmatrix} = \begin{bmatrix} \sum x \\ \sum xt \\ \sum xt^2 \\ \vdots \\ \sum xt^q \end{bmatrix} \quad (3.9)$$

Z této soustavy lineárních algebraických rovnic se najdou α_k známými metodami numerické matematiky. Bez počítače nebo alespoň programovatelné minikalkulačky je to ovšem úloha zdlouhavá již tehdy, je-li počet hledaných parametrů větší než tři. Řešení soustavy se však může usnadnit, volíme-li účelně funkce $g^{(k)}(t)$, aby se matice soustavy zjednodušila, a to tak, aby se v ní objevilo co nejvíce nulových prvků.

Maximálního zjednodušení dosáhneme, jestliže nenulovými zůstanou pouze prvky na hlavní diagonále a všechny ostatní budou nulové. K tomu je třeba, aby zvolená soustava funkcí $p^{(k)}(t)$ splňovala podmínky ortogonálnosti

$$\sum_{i=1}^n p_i^{(k)}(t) \cdot p_i^{(m)}(t) = 0, \quad k \neq m, \quad (3.10)$$

$$\neq 0, \quad k = m.$$

Pak se jednotlivé parametry α_k vypočtou jednotlivě, každý nezávisle na ostatních z jednoduchých vztahů

$$\alpha_k = \frac{\sum_{i=1}^n x_i \cdot p_i^{(k)}}{\sum_{i=1}^n [p_i^{(k)}]^2}. \quad (3.11)$$

Podrobné tabulky ortogonálních mnohočlenů lze nalézt v publikacích /33/ a /36/.

Vyhodnocování parametrů podle (3.11) je velmi přehledné a rychlé pomocí formuláře uvedeného na tabulce 2, kde se jednotlivá políčka vyplňují podle předpisu. Funkční hodnoty mnohočlenů

se opíší do horní třetiny tabulky 2. V prostředním pásu probíhá výpočet odhadu parametrů (po sloupec 2) a dále pak výpočet tzv. součtů čtverců, které slouží k testování významnosti vypočtených parametrů (podrobnosti o tom viz /32/). Ve spodní třetině formuláře se počítají vyrovnané hodnoty x_i^* pomocí těch členů sou-

TABULKA
VYROVNÁNÍ ZÁVISLOSTI ORTOGONÁLNÍMI FUNKCEMI

P _{ni} ^(k)	i							α _k
	1	2	3	4	5	6	7	
								α ₀
k	(1)							
	(2)							
	(3)							
	(4)							
	(5)							
	(6)							

$\hat{1} = \sum_{i=1}^n x_i$ resp. $\sum_{i=1}^n x_i^{(k)}$, resp. $\sum_{i=1}^n x_i^2$
 $\hat{2} = n$ pro x_i , resp. $\sum_{i=1}^n (p_{ni}^{(k)})^2$, pro $x_i p_{ni}^{(k)}$, resp. 1 pro x_i^2
 $\hat{3} = \hat{1} : \hat{2} = \bar{x}_k$ $\hat{6} = \hat{4} : \hat{5} =$ odhad složky rozptylu
 $\hat{4} = \hat{1} \cdot \hat{3} =$ součet čtverců $\hat{7} = \hat{6} : s^2 = F_p(k)$
 $\hat{5} =$ počet stupňů volnosti

								$\hat{1}$	$\hat{2}$	$\hat{3}$	$\hat{4}$	$\hat{5}$	$\hat{6}$	$\hat{7}$	F _{0,05}	F _{0,01}
x _i											= (K)					
x _i · p _{ni} ^(k)	(1)															
	(2)															
	(3)															
	(4)															
	(5)															
	(6)															
x _i ²											= Σ $\hat{4}$					
α ₀																
α _k · p _{ni} ^(k)	(1)															
	(2)															
	(3)															
	(4)															
	(5)															
	(6)															
x _i [*]																

\rightarrow = Σ $\hat{4}$
 s^2
 s_1^2
 s_2^2

stavy funkcí $p^{(k)}(t)$, které se prokázaly jako významné (obvykle jsou to členy s nejnižšími hodnotami indexu k).

9. ZÁVĚR.

Teorie pravděpodobnosti a matematická statistika poskytují řadu postupů pro vyhodnocování experimentálních údajů. Přitom mnohé z nich jsou aplikačně velmi výhodné pro svou jednoduchost a nenáročnost na technické prostředky. Je však třeba znát podmínky, za nichž byly tyto úsporné algoritmy odvozeny, jakož i meze jejich použitelnosti. V tomto příspěvku je ukázáno několik metodických přístupů a konkrétních návrhů, které splňují zmíněné požadavky a zaslouhují si proto pozornost širokého okruhu experimentátorů. Intenzivní využívání těchto metod by přispělo ke kvalitě jedné z důležitých etap každé experimentální práce, jíž je bezpochyby vyhodnocování experimentálních dat.

LITERATURA

- 1 Hátle, J.-Likeš, J.: Základy počtu pravděpodobnosti a matematické statistiky. Praha, SNTL 1974.
- 2 Mitropolskij, A.K.: Technika statističeskuch vyčislenij. Moskva, Nauka-FML, 1971.
- 3 ČSN 01 0250, Statistické metody v průmyslové praxi. Všeobecné základy, Praha, VÚNM 1975; ČSN 01 0252, Statistické metody v průmyslové praxi II-Závislosti mezi náhodnými veličinami-korelace a regrese. Praha, VÚNM 1977; ČSN 01 0253, Statistické metody v průmyslové praxi III-Základní neparametrické metody. Praha, VÚNM 1977.
- 4 Tukey, J.W.: On the comparative anatomy of transformations. Ann.Math.Statist. 28(1957), 602-632.
- 5 Box, G.E.P.-Cox, D.R.: An Analysis of transformations. J.Royal Statist.Soc.B 26(1964), 211-252 (with discussion).
- 6 Hinkley, D.: On power transformation to symmetry. Biometrika 62(1975), 101-111.
- 7 Dolby, J.L.: A quick method for choosing a transformation. Technometrics 5(1963), 317-326.
- 8 Hinkley, D.: On quick choice of power transformation. Appl. Statist. 26(1977), 67-69.
- 9 Johnson, N.L.: System of frequency curves generated by methods of translation. Biometrika 36(1949), 149-176; 297-304.
- 10 Sarhan, A.E.-Greenberg, B.G.: Contributions to order statistics. New York, J.Wiley (1962).
- 11 David, H.A.: Order Statistics. New York, J.Wiley (1970).
- 12 Lloyd, E.H.: Least-squares estimation of location and scale parameters using order statistics. Biometrika 39(1952), 88-95.
- 13 Chernoff, H.-Lieberman, G.J.: Use of normal probability paper. J.Amer.Statist.Assoc. 49(1954), 778-785.
- 14 Chernoff, H.-Lieberman, G.J.: The use of generalized probability paper for continuous distribution. Ann.Math.Statist. 27(1956), 806-818.
- 15 Kao, J.H.K.: A graphical estimation of mixed Weibull parameters in life testing of electron tubes. Technometrics 1, (1959), 389-407.
- 16 Daniel, C.: Use of half-normal plots in interpreting factorial two-level experiments. Technometrics 1(1959), 311-341.
- 17 Kimball, B.F.: On the choice of plotting positions on probability paper. J.Amer.Statist.Assoc. 55(1960), 546-560.
- 18 Wilk, M.B.-Gnanadesikan, R.-Huyett, M.J.: Probability plots for the gamma distribution. Technometrics 4(1962), 1-20.

- 19 Nelson, L.S.: Weibull probability paper. Industr. Quality Control 23(1966-67), 452-53.
- 20 McCool, J.I.: Inference from the third failure in a sample of 30 from the Weibull distribution. Industr. Quality Control 23(1966-67), 109-114.
- 21 Wilk, M.B.-Gnanadesikan, R.: Probability plotting methods for the analysis of data. Biometrika 55(1968), 1-17.
- 22 Healy, M.J.R.: Multivariate normal plotting. Appl. Statist. 17 (1968), 157-161.
- 23 Nelson, W.: Theory and application of hazard plotting for censored failure data. Technometrics 14(1972), 945-966.
- 24 Barnett, V.: Probability plotting methods and order statistics. Appl. Statist. 24(1975), 95-108.
- 25 Barnett, V.: Convenient probability plotting positions for the normal distribution. Appl. Statist. 25(1976), 47-50.
- 26 Kropáč, O.: Pravděpodobnostní papíry v experimentálním výzkumu. Zpravodaj VZLÚ 1(127), (1978), 3-45.
- 27 Kropáč, O.: O jednom způsobu zobrazení sdružené empirické distribuční funkce v pravděpodobnostních papírech. Zpravodaj VZLÚ 4(106), (1974), 5-13.
- 28 Kropáč, O.: Grafické zobrazení sdružených rozdělení pravděpodobnosti. Sborník přednášek STAGUAREL 80. Praha, Dům techniky ČSVTS (1980), 142-149.
- 29 Eggermayer, F.-Janeček, Z.: Grafická korelační analýza. Praha, SNTL (1968).
- 30 Dolby, J.L.: Graphical procedure for fitting the best line to a set of points. Technometrics 2(1960), 477-481.
- 31 Kropáč, O.: Vyrovnání experimentálních údajů ortogonálními soustavami funkcí s použitím statistických testů významnosti. Zpráva VZLÚ č.Z-5 (1964).
- 32 Fisher, R.S.-Yates, F.: Statistical Tables for Biological, Agricultural and Medical Research. Edinburgh, Oliver & Boyd (4. vydání 1953).
- 33 Bolšev, L.N.-Smirnov, N.V.: Tabлицы математической статистики. Москва, Наука (1965).
- 34 Grabar, L.P. Таблицы полиномов Чебышева, ортонормированных на системе равноотстоящих точек, Москва, Вычисл. центр АН СССР (1965).
- 35 Likeš, J.-Laga, J.: Základní statistické tabulky. Praha, SNTL (1978).