

**Univerzita Komenského v Bratislave**  
Prírodovedecká fakulta  
Katedra aplikovanej a environmentálnej geofyziky

---

**Vojtech Gajdos**

**Stručný prehľad štatistiky pre geológov**

Bratislava 2004

## 1. Úvod

Štatistika je veda, ktorá sa zaoberá hodnotením hromadných javov. Jednotlivý prípad môže skúmať príslušná vedná disciplína a podať o ňom výpoveď. Napríklad ak skúmame vzorku horniny, môžeme zistiť jej mineralogické zloženie, popísať jednotlivé minerály, ktoré sú v nej prítomné, popísať štruktúru vzorky, jej magnetické vlastnosti, rádioaktivitu a pod. Ak však chceme porovnávať vzorky hornín navzájom, dostávame sa na pôdu štatistiky. Pri porovnávaní vzoriek hornín sa totiž dostávame na kvalitatívne inú úroveň poznania: môžeme usudzovať o širších súvislostiach medzi vzorkami hornín, pretože každá vzorka je súčasťou rozsiahlejšieho geologického telesa a cez poznanie jednotlivých vzoriek môžeme získať poznatky o celom geologickom telese z ktorého vzorka pochádza a ďalej, poznaním geologického telesa môžeme porovnávať navzájom jednotlivé geologické telesá a tak postupne pochopiť štruktúru a históriu väčších geologických celkov a celej geologickej histórie planéty. Štatistiku teda musíme použiť vtedy, keď chceme porovnávať navzájom viacero jednotlivých prípadov a tak sa dostať k poznaniu širších súvislostí medzi nimi.

Uvedený postup má však jednu ťažkosť. Nie je prakticky možné a často ani potrebné rozdeliť celé geologické teleso ktoré chceme preskúmať na vzorky a tieto preskúmať a určiť charakteristiky potrebné pre ďalšie štatistické spracovanie. Preto je ambíciou štatistiky na základe hodnotenia malej časti prvkov (vzoriek) zostaviť objektívnu charakteristiku celého (úplného) súboru prvkov (geologického telesa).

## 2. Základné pojmy

Uvažujme geologickú vrstvu. Táto vrstva má svoju hrúbku, hĺbku, mineralogické a petrografické zloženie, ak je to usadená vrstva, tak obsahuje zostatky určitej fauny a flóry, materiál vrstvy má rôzne fyzikálne a technologické vlastnosti a zrejme aj rad ďalších charakteristík.

Ak by sme chceli zostaviť úplnú charakteristiku takejto vrstvy, museli by sme analyzovať každý kubický centimeter vrstvy a pri niektorých vlastnostiach by ani táto objemová jednotka nestačila a pri iných by bola zbytočne podrobná. Je teda zrejmé, že úplnú charakteristiku nie je možné zostaviť a preto sa musíme uspokojiť s približnou charakteristikou vrstvy. Napriek tomu chceme, aby sa zostavené približné charakteristiky významne nelíšili od skutočných. Otázkou teda je, ako to urobiť. Odpoveďou je štatistika. Štatistika je metóda, pomocou ktorej je možné zostaviť charakteristiky našej hypotetickej vrstvy pri minimálnom počte vzoriek na ktorých budeme tieto charakteristiky študovať. Ak teda budeme dodržiavať postupy ktoré štatistika predpisuje tak charakteristiky ktoré získame, sa budú iba nevýznamne líšiť od tých skutočných.

Jav. Ako sme uviedli, štatistika neskúma jednotlivé javy, ale hromadné javy. Javom nazývame fakt, ktorý ako výsledok pokusu, hodnotenia, skúmania či merania môže, ale nemusí nastať. Napríklad:

- a) padnutie šestky pri hode hracou kockou
- b) výskyt určitého minerálu v sérii vzoriek.
- c) výskyt anomálie magnetického poľa na vymedzenej časti územia.

Výsledok je teda náhodný a miera náhodnosti závisí na množstve okolností, ktoré túto náhodnosť podmieňujú. Dobrým príkladom je šport, kde dosiahnutie želaného výsledku závisí, hlavne pri veľkej vyrovnanosti aktérov, často na mnohých, zdanlivo nepodstatných

okolnostiach a ich vzájomnej súhre (hovoríme tomu synergia). Náhodné javy sa tiež označujú ako stochastické.

Pravdepodobnosť javu. Je to určité číslo, ktorým kvantitatívne hodnotíme stupeň toho, že jav nastane. Toto číslo je tým väčšie, čím je väčšia možnosť toho, že jav nastane.

Štatistická jednotka. V úvode spomínaná horninová vzorka predstavuje príklad štatistickej jednotky (definovaná časť z celého súboru). Podobnými štatistickými jednotkami môžu byť osoby, veci, udalosti a pod.

Znak. Každá štatistická jednotka má aspoň jednu charakteristickú vlastnosť, pre ktorú je zaradená do štatistického skúmania (farba, váha, druh, ...). Táto vlastnosť sa nazýva štatistický znak (niekedy sa nazýva aj meraný znak a niekedy aj štatistický prvok). Prvky štatistického súboru musia mať niektoré vlastnosti, ktoré sú všetkým prvkom spoločné a vyjadrujú hľadisko vecné, časové a priestorové. Druhá skupina spoločných vlastností sú vlastnosti vyšetřované. Všetky prvky musia mať aspoň jednu vyšetřovanú vlastnosť, ktorá sa nazýva **meraný znak** (alebo len **znak**). Znaky sa podľa svojej povahy môžu deliť na kvantitatívne (sú merateľné, dajú sa vyjadriť číselne), a kvalitatívne (dajú sa vyjadriť popisom, napr. farba).

Štatistický súbor. Uvedieme príklad: pri meraní vzoriek hornín narezaných z jedného bloku andezitu sa zistili tieto hodnoty magnetickej susceptibility ( $\kappa \times 10^{-6}$  cgs): 350,7; 350; 351,4; 350; 351,1; 350,9; 351,2; 351; 351,1; 351,3; 351,1; 351,2; 351; 351; 350,9; 350,8. Bolo teda vykonané hromadné pozorovanie, inými slovami, vykonané bolo väčšie množstvo meraní vzoriek, aby sa posúdila hodnota susceptibility bloku horniny. Každá vzorka (štatistická jednotka) je zastúpená jedinou hodnotou (znakom). Skupina znakov tvorí **štatistický súbor** (alebo len **súbor**). Súbor, ktorý zahŕňa všetky znaky vyhovujúce daným vecným, časovým a priestorovým kritériám sa nazýva **základný súbor**. Ako sme uviedli, je prakticky nemožné pracovať s celým základným súborom. To znamená, že súbor s ktorým pracujeme, je **výber** zo základného súboru.

Náhodný výber. Len málokedy máme možnosť zostaviť reprezentatívny výber zo základného súboru. Je to preto, že tento základný súbor nepoznáme, nepoznáme jeho zákonitosti. Preto sa pri charakterizovaní súboru používa tzv. **náhodný výber**, ktorého zostavenie je založené na podmienke, aby každý znak základného súboru mal rovnakú nezávislú pravdepodobnosť (rovnakú nepodmienenú šancu), že bude do výberu zahrnutý.

Náhodná veličina. Znak je veličina **premenná**. Ak nemôžeme na základe určitej zákonitosti dopredu stanoviť hodnotu tejto veličiny, je výskyt každej jej jednotlivéj hodnoty dôsledkom náhodného pôsobenia známych, aj neznámych vplyvov. V takomto prípade nadobúda znak konkrétnu veličinu náhodne. Túto veličinu nazývame **náhodná veličina**.

Spojité veličina. Pod spojitou veličinou rozumieme takú náhodnú veličinu, ktorá môže nadobudnúť v určitom intervale akúkoľvek hodnotu (napr. 2,673). Príkladom súboru, pri ktorom znaky prvkov sú spojité veličiny je napr.: objem neopracovanej vzorky, prirodzená remanentná magnetizácia, hmotnosť konkrétneho človeka, jeho výška a pod.

Diskrétna (nespojité) veličina. Pod diskretnou veličinou rozumieme takú náhodnú veličinu, ktorá môže nadobudnúť iba určité hodnoty. Napr. ak hádzeme hracou kockou, môžeme dostať iba celé čísla od 1 do 6, t.j. nemôžeme dostať číslo 2,3; 4,1 a pod.

Rozsah štatistického súboru. Počet všetkých znakov štatistického súboru sa nazýva rozsah a značí sa písmenom **n**.

Triedy. V prípade, že musíme pracovať zo štatistickým súborom veľkého rozsahu, je výhodné rozdeliť si tento súbor na menšie rovnaké časti - **triedy**, alebo **skupiny**.

Početnosť. Predpokladajme, že istý pokus bol vykonaný n-krát, pričom výsledok A sme dostali p-krát (napríklad pri hádzaní kockou sme hádzali 54 krát a pritom trojka padla 8 krát). Potom číslo p nazývame **početnosťou** výsledku A. Pojem početnosti používame najčastejšie pri súboroch rozdelených na triedy. Potom jedna trieda bude zastúpená určitou strednou hodnotou - **triednym znakom** a **triednou početnosťou** – počtom prvkov v tejto triede.

Kumulatívna početnosť. Majme veľký súbor, rozdelený na triedy, ktorých početnosti sú:  $n_1, n_2, \dots, n_j, \dots, n_r$ . Potom členov postupnosti  $n_1; n_1 + n_2; \dots; \sum_{i=1}^j n_i; \dots; \sum_{i=1}^r n_i$  nazývame **kumulatívne početnosti**.

Rozdelenie početností štatistického súboru. Po rozdelení štatistického súboru do tried zistíme, že v každej triede nie je rovnaká početnosť. O vzťahu medzi hodnotami súboru a ich početnosťami nás informuje rozdelenie početností štatistického súboru.

Štatistické skúmanie. Štatistické skúmanie hromadných javov sa vo všeobecnosti rozdeľuje na tri etapy:

- štatistické zisťovanie
- štatistické spracovanie
- štatistický rozbor.

### 3. Grafické zobrazenie štatistického súboru a jeho prvkov

Spracovanie výsledkov merania začíname zvyčajne ich grafickým zobrazením. Zobrazenie súboru nám umožňuje získať si predstavu o jeho tvare, naznačiť niektoré zákonitosti, ukázať cestu ďalšieho numerického spracovania.

Štatistické tabuľky. Získané štatistické dáta sa triedia a zapisujú v stručnej a prehľadnej podobe do tabuliek. **Štatistické tabuľky** sú dôležitým oznamovaním štatistických prvkov a preto sú normalizované. Tvar tabuľky je uvedený na obr.1.

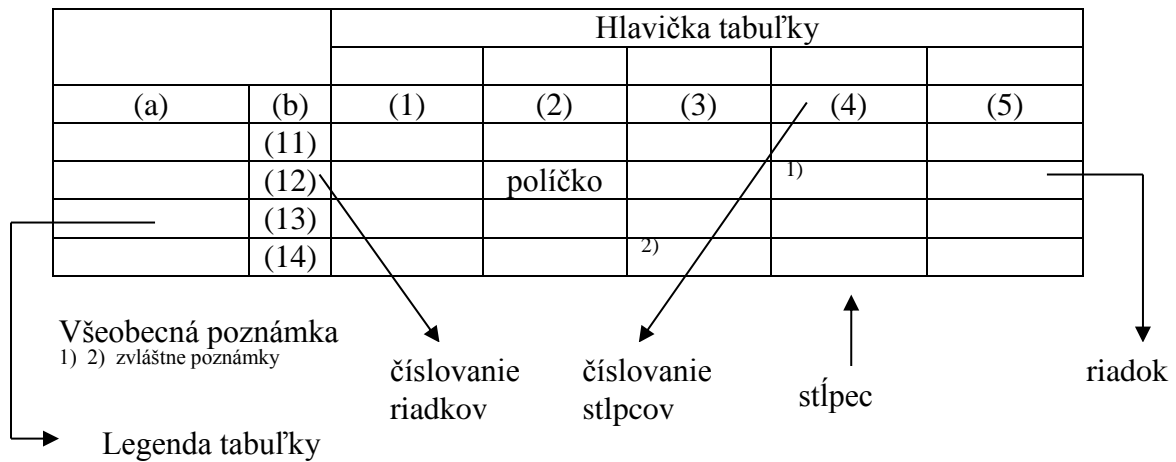
Grafy. Graf je zobrazením vzájomných vzťahov dvoch, alebo viac premenných veličín pomocou prehľadných symbolov: bodov, úsečiek, čiar, štvorcov, obdĺžnikov, kruhov a pod. Graf je najjednoduchším zobrazením závislosti

$$y = f(x) \tag{1}$$

Aby bol graf jednoduchý a dobre čitateľný, musí mať stručný názov a presné označenie stupnice. Zostrojuje sa najčastejšie pomocou grafických papierov (milimetrový, semilogaritmický, logaritmický a pod.). Dôležitú pozornosť pri realizácii grafu je potrebné

venovať modulu stupnice. V opačnom prípade by mohol byť skreslený. Na horizontálnu os (abscisa) nanášame nezávislú premennú, na vertikálnu os (ordináta) závisle premennú.

### Nadpis tabuľky



**Obr.1.** Tvar základnej štatistickej tabuľky

Modul stupnice, t.j. vzdialenosť medzi dvomi celými susednými hodnotami vynášanej veličiny, vyjadrenú v cm, vypočítame podľa vzťahu:

$$\varepsilon = \frac{L}{R} = \frac{L}{x_{\max} - x_{\min}} \quad (2)$$

$$R = x_{\max} - x_{\min} \quad (3)$$

kde  $\varepsilon$  je modul v cm;  $L$  je požadovaná dĺžka stupnice v cm;  $R$  je variačné rozpätie;  $x_{\max}$  je maximálna hodnota vynášanej veličiny;  $x_{\min}$  je minimálna hodnota vynášanej veličiny.

Pri zostrojovaní grafu početností, je veľkosť intervalu zhodná s veľkosťou tried, na ktoré bol súbor rozdelený. Šírku triedneho intervalu môžeme určiť podľa vzťahu:

$$h = 0,08 R \quad (4)$$

alebo

$$h < \frac{R}{12} < 2 h \quad (5)$$

kde  $R$  je variačné rozpätie.

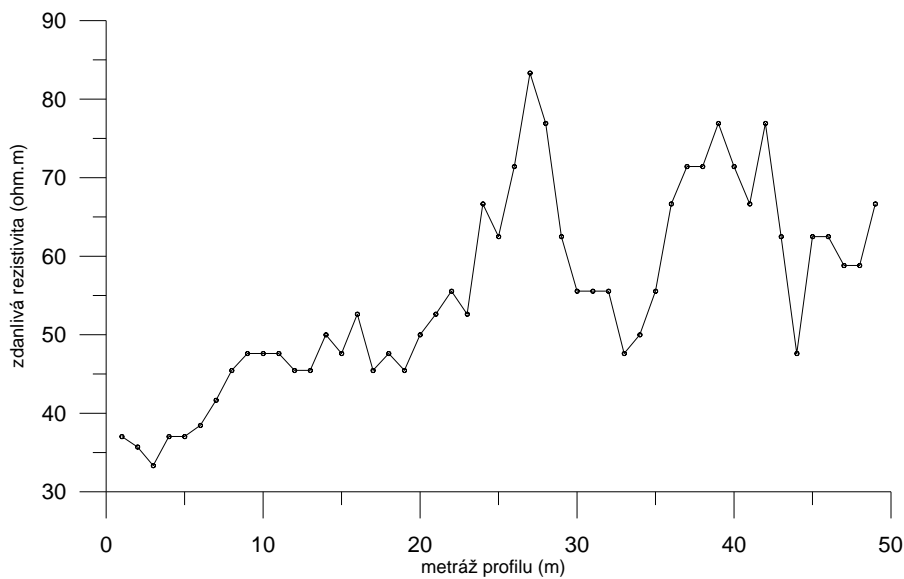
Veľkosť  $h$  možno tiež určiť podľa vzťahu:

$$h = \frac{X_{\max} - X_{\min}}{1 + 3,2 \log n} = \frac{R}{1 + 3,2 \log n} \quad (6)$$

kde n je rozsah súboru. Vzťah (6) možno využiť na určenie počtu tried

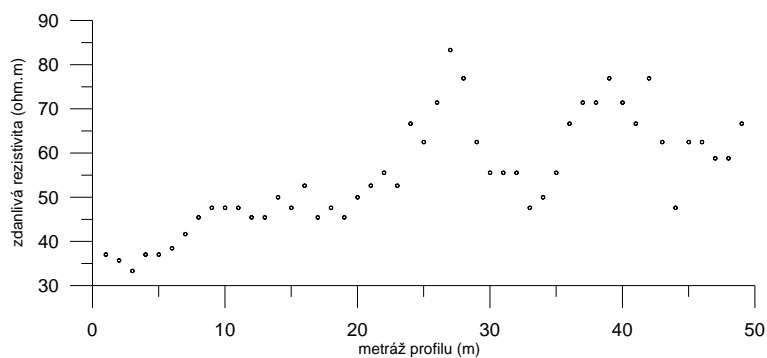
$$k = 1 + 3,2 \log n \quad (6a)$$

Spojnicový graf. Najčastejšou formou grafu je spojnicový graf. Jednotlivé hodnoty sa zakresľujú pomocou bodov. Body spájame priamymi, alebo krivými čiarami podľa povahy zobrazovaného vzťahu (obr.2).



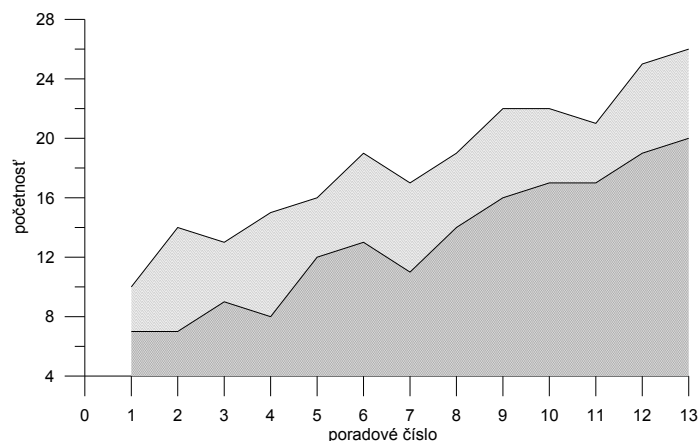
**Obr.2.** Spojnicový graf

Bodový graf. V grafoch, ktoré zobrazujú korelačné vzťahy sa jednotlivé hodnoty – body , nespájajú. Dostávame tak bodový graf (obr.3).



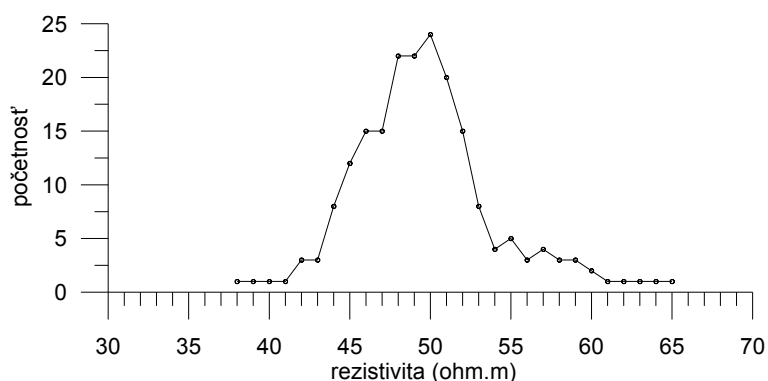
**Obr.3.** Bodový graf

Povrchový graf. Niekedy pre lepšiu názornosť je vhodné plochy, ktoré graf vymedzuje, vyšrafovať. Robí sa to hlavne v tom prípade, ak na jednu plochu kreslíme dva grafy. Takýto graf sa nazýva tiež povrchový (obr.4).



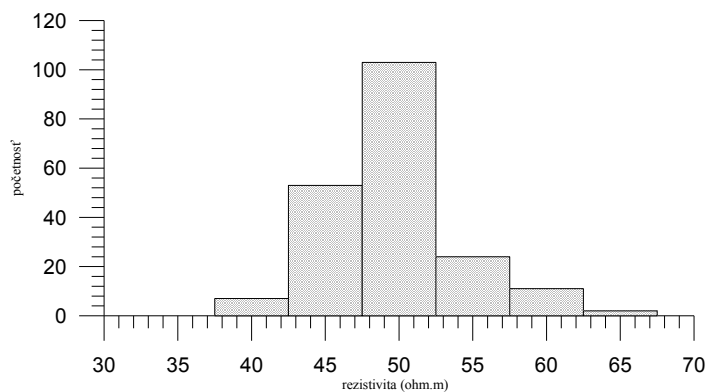
**Obr.4.** Povrchový graf

Polygón početností. Spojnicovým grafom sa tiež graficky znázorňuje rozloženie diskkrétnej veličiny. Takýto graf nazývame mnohoúhelník (polygón) početností. Jednotlivé body polygónu početností sa spájajú priamou čiarou. Polygónom početností sa zobrazujú len súbory rozdelené na triedy. Jednej triede potom odpovedá jeden bod (obr.5).



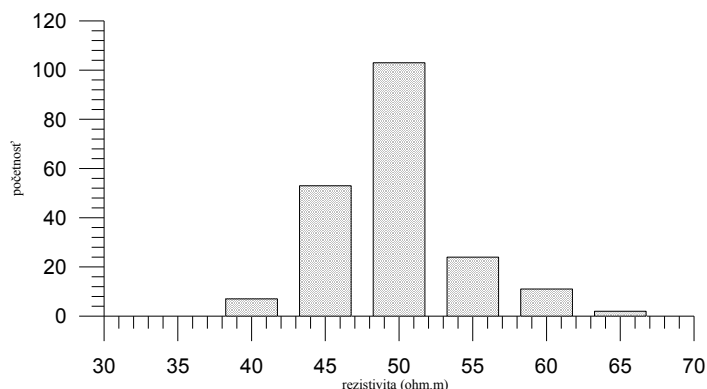
**Obr.5.** Polygón početností.

Histogram. Rozloženie početností sa najčastejšie zobrazuje pomocou histogramu. Na rozdiel od polygónu početností, sa v histograme početnosť zobrazuje úsečkou o dĺžke triedneho intervalu. Okraje úsečky ohraničíme zvislými čiarami. Dostaneme tak stĺpčeky. Potom šírka stĺpčeka je rovná veľkosti triedneho intervalu, jeho výška je rovná početnosti príslušného triedneho intervalu. Netreba zdôrazňovať, že všetky stĺpčeky histogramu musia mať rovnakú šírku (obr.6).



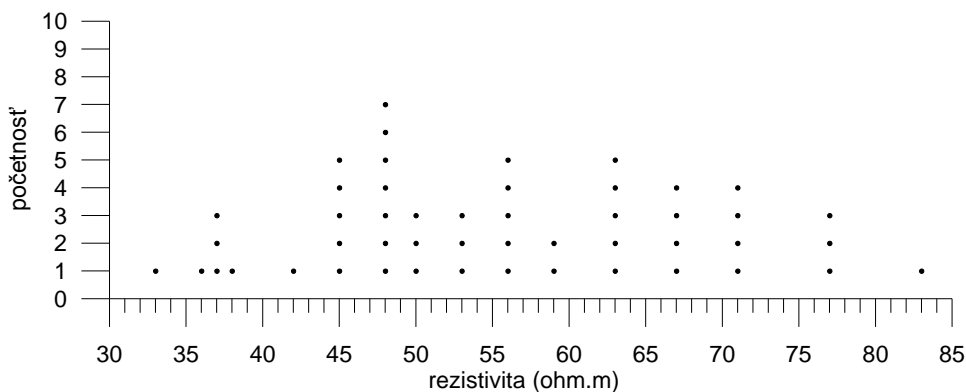
**Obr.6.** Histogram

Stĺpčekový diagram. Má rovnaký význam ako histogram s tým rozdielom, že úsečka nemá dĺžku celého intervalu, ale len jeho časti. Šírka jednotlivých stĺpčekov však musí byť konštantná (obr.7).



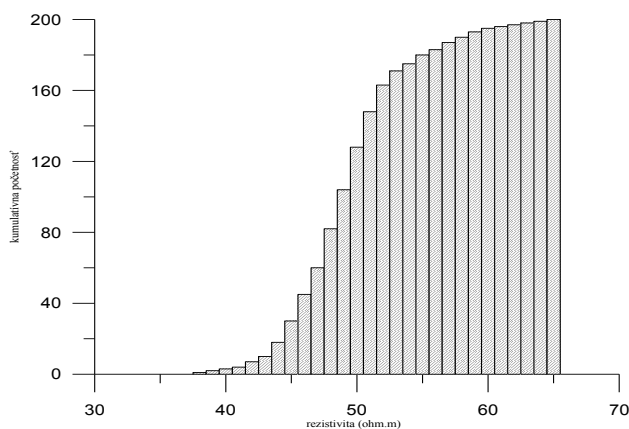
**Obr.7.** Stĺpčekový diagram

Bodový diagram. Má rovnaký význam ako histogram. Pri tomto diagrame sa však početnosť zobrazuje pomocou bodov, ktoré sa nespájajú [3]. Každý prvok má v bodovom diagrame svoj bod v príslušnom triednom intervale v ktorom sa nachádza. Bodový diagram sa používa vtedy, keď zobrazovaný súbor zaberá veľký interval a má malý rozsah. Pri tom chceme, aby triedne intervaly zostali malé (obr.8).



**Obr.8.** Bodový diagram

Diagram kumulatívnej početnosti. Diagram kumulatívnej početnosti získame podobne ako histogram s tým rozdielom, že pri diagrame kumulatívnej početnosti výška jednotlivých stĺpčekov odpovedá príslušným hodnotám kumulatívnych početností (obr.9).



**Obr.9.** Diagram kumulatívnej početnosti



## 4. Štatistické charakteristiky

Ako sme uviedli, úlohou štatistiky je, pre daný súbor hodnôt nájsť okrem iného určité číselné charakteristiky, pomocou ktorých možno súbor popísať, analyzovať t.j. vyťažiť s neho všetky dostupné informácie.

Takýchto číselných charakteristík sa používa niekoľko druhov a každá informuje o určitej špecifickej vlastnosti súboru. Preto je potrebné ich výber previesť účelne.

### 4.1. Stredné hodnoty

Vždy pri štatistickej práci sa hľadá jedna taká hodnota, ktorá by čo najlepšie charakterizovala študovaný súbor. Pre ideálny súbor je to aritmetický priemer. Avšak v praxi sa málo kedy stretáme s ideálnym súborom. Obyčajne zaberá študovaný súbor široký interval hodnôt s nerovnomerným rozdelením početností, často sa v súbore vyskytujú extrémne hodnoty. Pretože aritmetický priemer takejto súbory dostatočne nevystihuje, používajú sa tiež iné stredné hodnoty ako medián, modus a pod.

Aritmetický priemer -  $\bar{x}$ . Aritmetický priemer počítame podľa nasledovného vzťahu

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (7)$$

kde  $n$  je rozsah súboru a  $x_i$  je  $i$ -ty prvok súboru. V prípade, že máme súbor rozdelený na triedy, počítame aritmetický priemer podľa vzťahu

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j \cdot n_j \quad (8)$$

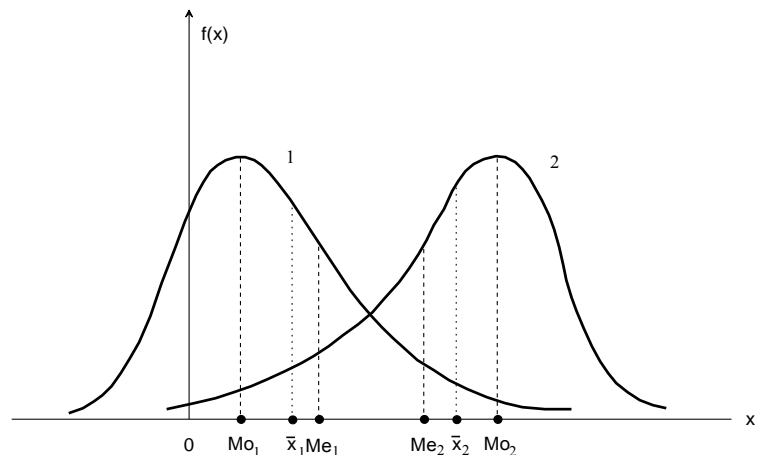
kde  $x_j$  je hodnota zastupujúca  $j$ -ty interval,  $n_j$  je počet prvkov  $j$ -tej triede, t.j. početnosť  $j$ -tej triedy.

Medián -  $\tilde{x}$ , Me. Ak študovaný súbor zoradíme podľa veľkosti, potom hodnota prostredného prvku nám udáva medián. V prípade párneho počtu prvkov, je to aritmetický priemer z dvoch prostredných prvkov. Výhoda mediánu oproti aritmetickému priemeru je v tom, že vplyv extrémnych hodnôt je v prípade mediánu vylúčený. Takto medián lepšie reprezentuje súbor a citlivejšie sa prikláňa k charakteristickej časti súboru. Pri tom určenie mediánu je rýchle, jednoduché, takže pravdepodobnosť omylu je menšia. V prípade normálneho rozdelenia početností súboru, je medián zhodný s aritmetickým priemerom. Medián je vhodné použiť na charakterizovanie súboru s  $n < 10$ .

n	$x_i$	$x_i$
1	350,7	350,0
2	350,0	350,0
3	351,4	350,7
4	350,0	350,8
5	351,1	350,9
6	350,9	350,9
7	351,2	351,0
8	351,0	351,0
9	351,1	351,0
10	351,3	351,1
11	351,1	351,1
12	351,2	351,1
13	351,0	351,2
14	351,0	351,2
15	350,9	351,3
16	350,8	351,4
suma =	5614,7	5614,7
	a.p. =	350,92
	medián =	351,00
	módus =	351,05
	$Q_{25}$ =	350,35
	$Q_{75}$ =	351,25

**Modus** – Mo. Modus je hodnota, ktorá sa v súbore najčastejšie objavuje – opakuje. Módus má zmysel použiť iba vtedy, keď dochádza k opakovaniu hodnôt.

**Kvantily** –  $Q_i, Q_{100-i}$ . Niekedy potrebujeme poznať polohu a rozloženie prvkov v súbore podrobnejšie, ako to ukazujú priemer a stredné hodnoty. Napríklad potrebujeme poznať úroveň extrémnych hodnôt. V takomto prípade nám informáciu môžu poskytnúť kvantily, čo je spoločný názov pre celú sústavu ukazovateľov polohy.



**Obr.10.** Poloha aritmetického priemeru, mediánu a módu pri asymetrickom štatistickom súbore

Ak máme súbor usporiadaný vo vzostupnom rade hodnôt, potom vo všeobecnosti označujeme symbolom  $Q_i$  kvantil v dolnej polovici súboru a symbolom  $Q_{100-i}$  kvantil v hornej polovici súboru. Index  $i$  ( $\leq 50$ ) znamená percento z celkového počtu hodnôt menších ako  $Q_{2i}$ , resp. väčších ako  $Q_{100-2i}$ . Tak napríklad kvantily ( $Q_{25}; Q_{75}$ ) určíme nasledovne:  $Q_{25}$  je medián dolnej polovice vzostupne usporiadaných hodnôt a  $Q_{75}$  je medián hornej polovice vzostupne usporiadaných hodnôt súboru. Podobne určíme decily ( $Q_{10}$  a  $Q_{90}$ ) percentily ( $Q_1$  a  $Q_{99}$ ) a pod. Medián je potom vlastne kvantil  $Q_{50}$ . Vysoké alebo nízke percentily sa používajú iba pri dostatočne veľkom súbore.

*Poznámka.* V teórii pravdepodobnosti sa vo vzťahoch pre stredné hodnoty, ako aj v iných vzťahoch, používa pojem pravdepodobnosť. Napr. pre aritmetický priemer možno s použitím pravdepodobnosti napísať

$$\bar{x} = \sum_{i=1}^n x_i p_i \quad (9)$$

Hodnota  $p_i$  značí pravdepodobnosť objavenia sa hodnoty  $x_i$  v súbore. Pravdepodobnosť sa určuje nasledovným spôsobom: pravdepodobnosť  $p_i$  javu  $A$  vypočítame ako pomer počtu prípadov, v ktorých jav  $A$  nastal ( $m_i$ ) a počtu celkového množstva prípadov ( $n$ )

$$p_i = \frac{m_i}{n} \quad (10)$$

*Prakticky:* v súbore s  $n = 100$ , rozdelenom na triedy, v triede 300 – 399 sa vyskytuje 30 hodnôt ( $m_i = 30$ ). Potom pravdepodobnosť výskytu ktorejkoľvek hodnoty z intervalu 300 – 399 v študovanom súbore je

$$p_i = \frac{m_i}{n} = \frac{30}{100} = 0,3$$

Podľa teórie pravdepodobnosti platí vzťah

$$\sum_{i=1}^n p_i = 1 \quad (11)$$

Je zrejmé, že ak súbor nie je rozdelený na triedy, potom pravdepodobnosť výskytu ktorejkoľvek hodnoty (ak sa vyskytuje iba raz) je  $1/n$ .

## 4.2 Charakteristiky rozptylu

Disperzia –  $s^2$ . Disperzia je definovaná ako priemer štvorcov odchýliek prvkov súboru od jeho aritmetického priemeru

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (12)$$

Alebo

$$s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad (13)$$

V prípade, že súbor je rozdelené na triedy platí

$$s^2 = \frac{1}{n} \sum_{j=1}^n n_j x_j^2 - \bar{x}^2 \quad (14)$$

Symbolika je rovnaká ako pri aritmetickom priemere.

Smerodajná odchýlka –  $s$  (tiež stredná kvadratická odchýlka, alebo niekedy aj štandard).

$$s = \sqrt{s^2} \quad (15)$$

Variačné rozpätie. Pre porovnanie rozptylov je niekedy výhodné vylúčiť vplyv jednotiek a preto sa používa ďalšia charakteristiky rozptýlenia – variačný koeficient, ktorý sa udáva v % a počíta podľa vzťahu

$$V(\%) = 100 \times \frac{s}{\bar{x}} \quad (16)$$

## 4.3 Iné charakteristiky

### Momenty

Pre opis základných vlastností súboru sa používa pojem momentu, podobne ako sa v mechanike používa moment inercie na opísanie rozloženia hmoty. V praxi sa najčastejšie používajú momenty dvoch typov: počiatkové a centrálné.

**Počiatočným momentom** s-tého rádu diskkrétnej veličiny  $X$  sa nazýva suma typu

$$\alpha_s[X] = \sum x_i^s p_i \quad (17)$$

Ako sme uviedli, aritmetický priemer možno vypočítať podľa vzťahu

$$\bar{x} = \sum_{i=1}^n x_i p_i \quad (18)$$

Ak poznáme oba vzťahy a porovnáваме ich vidíme, že aritmetický priemer je vlastne počiatočný moment s-tého rádu náhodnej veličiny  $X$  a preto sa nazýva aritmetický priemer s-tého stupňa tejto veličiny.

**Centrálny moment.** Najprv si osvetlíme pojem centrovanej náhodnej veličiny. Centrovanou náhodnou veličinou budeme nazývať odchýlku náhodnej veličiny  $X$ , od jej aritmetického priemeru

$$\dot{X} = x - \bar{x} \quad (19)$$

kde  $\dot{X}$  je centrovaná náhodná veličina.

Aritmetický priemer centrovanej náhodnej veličiny je rovný nule

$$\bar{X}[\dot{X}] = \bar{X}[x - \bar{x}] = \sum_{i=1}^n (x_i - \bar{x}) p_i = \sum_{i=1}^n x_i p_i - \bar{x} \sum_{i=1}^n p_i = \bar{x} - \bar{x} = 0$$

Nezabudnime, že  $\sum_{i=1}^n p_i = 1$  a  $\sum_{i=1}^n x_i p_i = \bar{x}$  (viď poznámku na str.10). Centrovanie náhodnej veličiny odpovedá prenosu počiatku súradníc do bodu, ktorý má hodnotu aritmetického priemeru.

Momenty centrovanej náhodnej veličiny sa nazývajú **centrálne momenty**. Sú analogické pojmu ťažisko v mechanike.

Centrálnym momentom s-tého rádu náhodnej veličiny  $X$  sa nazýva aritmetický priemer s-tého stupňa tejto centrovanej náhodnej veličiny

$$\mu_s[X] = \bar{X}[\dot{X}^s] = \bar{X}[(x - \bar{x})^s] \quad (20)$$

Pre diskrétnu veličinu sa s-tý centrálny moment vyjadruje sumou

$$\mu_s = \sum_{i=1}^n (x_i - \bar{x})^s p_i \quad (21)$$

Ako sme ukázali, prvý centrálny moment je rovný nule

$$\mu_1 = 0$$

Vypočítame si druhý centrálny moment

$$\mu_2 = \sum_{i=1}^n (x_i - \bar{x})^2 p_i = \sum_{i=1}^n x_i^2 p_i - 2\bar{x} \sum_{i=1}^n x_i p_i + \bar{x}^2 \sum_{i=1}^n p_i = \alpha_2 - 2\bar{x}^2 + \bar{x}^2 = \alpha_2 - \bar{x}^2 \quad (22)$$

Analogicky vypočítame tretí centrálny moment

$$\mu_3 = \sum_{i=1}^n (x_i - \bar{x})^3 p_i = \sum_{i=1}^n x_i^3 p_i - 3\bar{x} \sum_{i=1}^n x_i^2 p_i + 3\bar{x}^2 \sum_{i=1}^n x_i p_i - \bar{x}^3 \sum_{i=1}^n p_i = \alpha_3 - 3\alpha_2 \bar{x} - 2\bar{x}^2 \quad (23)$$

(zmysel hodnoty  $p_i$  je vysvetlený v poznámke na str.10).

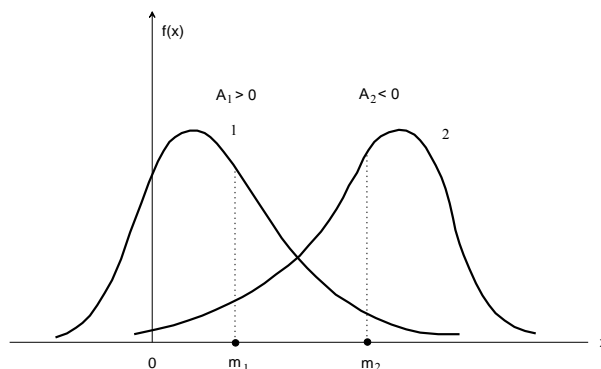
**Druhý centrálny moment** sa nazýva disperzia náhodnej veličiny

$$\mu_2 = s^2 \quad (24)$$

Ak použijeme analógiu s mechanikou, tak disperzia je vlastne moment inercie – udáva rozdelenie hmoty okolo ťažiska.

**Tretí centrálny moment** sa používa na charakteristiku symetrie resp. asymetrie súboru (vlastne každý nepárny moment sa môže použiť na charakteristiku asymetrie). Aby sme dostali bezrozmernú veličinu, delíme tretí centrálny moment treťou mocninou strednej kvadratickej odchýlky. Získanú hodnotu nazývame asymetria:

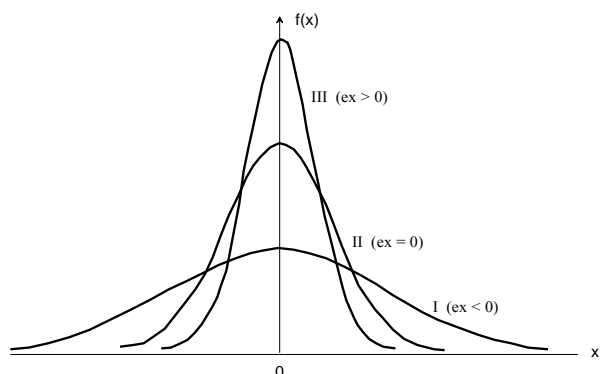
$$A = \frac{\mu_3}{s^3} \quad (25)$$



*Obr.11. Asymetria štatistického súboru*

**Štvrtý centrálny moment** sa používa na charakteristiku štiňlosti (špicatosti) súboru. Získanú veličinu nazývame exces

$$E = \frac{\mu_4}{s^4} - 3 \quad (26)$$



*Obr.12. Exces štatistického súboru*

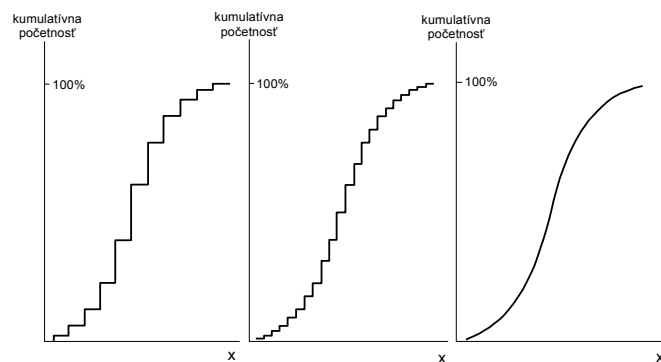
## 5. Štatistické rozdelenia

Ako sme uviedli, o vzťahu medzi hodnotami súboru a ich početnosťou nás informuje rozdelenie štatistického súboru.

Graficky zobrazujeme rozdelenie diskkrétnej veličiny pomocou polygónu početnosti a histogramom (obr.5 a obr.6), alebo diagramom kumulatívnej početnosti (obr.9).



**Obr.13.** Prechod od histogramu k frekvenčnej funkcii



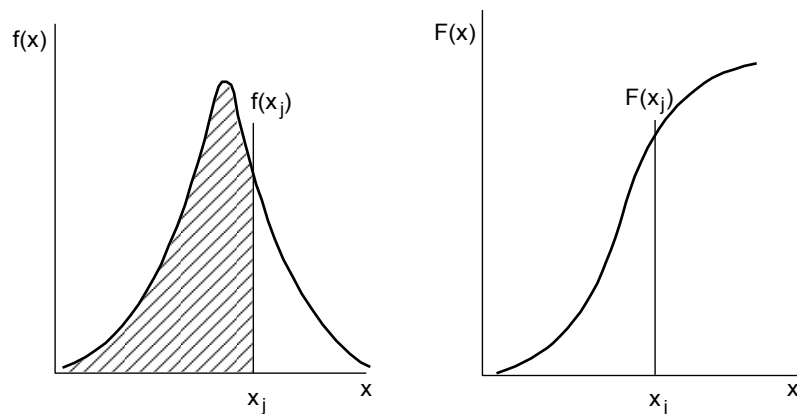
**Obr.14.** Prechod od kumulatívnej početnosti k distribučnej funkcii

Rozdelenie spojitkej veličiny sa zobrazuje pomocou krivky, ktorá sa nazýva **frekvenčná funkcia** –  $f(x)$ , (v teórii pravdepodobnosti sa nazýva hustota pravdepodobnosti alebo hustota rozdelenia, či diferenciálna funkcia rozdelenia). Frekvenčná funkcia je deriváciou tzv. **distribučnej funkcie** –  $F(x)$ . Frekvenčná a distribučná funkcia spojitých náhodných veličín majú niektoré dôležité vlastnosti, najmä tieto:

- obidve funkcie nikdy neklesnú pod vodorovnú os (t.j. nikdy nie sú záporné)
- frekvenčná krivka vymedzuje nad vodorovnou osou plochu, ktorej veľkosť má hodnotu 1
- distribučná funkcia zaberá vždy celý interval od 0 po 1. Má v tomto intervale neklesajúci priebeh.
- hodnota distribučnej funkcie v bode  $x_j$  –  $F(x_j)$  (obr.15), predstavuje plochu, ktorú ohraničuje frekvenčná funkcia nad vodorovnou osou od svojho počiatku až k poradnici  $f(x_j)$  v bode  $x_j$  (obr.15).

Obe funkcie majú veľký význam pre teóriu matematickej štatistiky, pretože slúžia na odvodenie a použitie celého radu veľmi dôležitých metód. Každý súbor má svoje rozdelenie, svoju frekvenčnú a distribučnú krivku. Každý súbor sa dá týmito tromi charakteristikami presne popísať; pre každý súbor sú charakteristické. Ak raz pre určitú náhodnú veličinu určíme jej rozdelenie, pri budúcom meraní tejto veličiny bude stačiť menší súbor hodnôt na to, aby sme určili jej charakteristiky ( $\bar{x}$ ,  $s$ ,  $A$ ,  $E$ , ...). V praxi sa zistilo, že súbory náhodnej

veľičiny vytvárajú niekoľko typov rozdelení. Pre tieto typy sa potom našli teoretické modely – teoretické rozdelenia a pomocou týchto modelov sa určili ich charakteristiky. V ďalšom si uvedieme niektoré teoretické rozdelenia, ktoré sa používajú v praxi FVH.



**Obr.15.** Frekvenčná a distribučná funkcia

Normálne rozdelenie (tiež Gaussovo rozdelenie). Základným teoretickým rozdelením je normálne rozdelenie. V praxi sa vyskytuje najčastejšie. Hlavná jeho výhoda je v tom, že je hraničným rozdelením, ku ktorému sa približujú ostatné rozdelenia pri najčastejšie prebiehajúcich typických podmienkach, t.j. ak máme dostatočne veľký súbor, tak v každom prípade, ak nemá priamo normálne rozdelenie, bude sa mu veľmi približovať.

Frekvenčná krivka normálneho rozdelenia je vyjadrená vzťahom

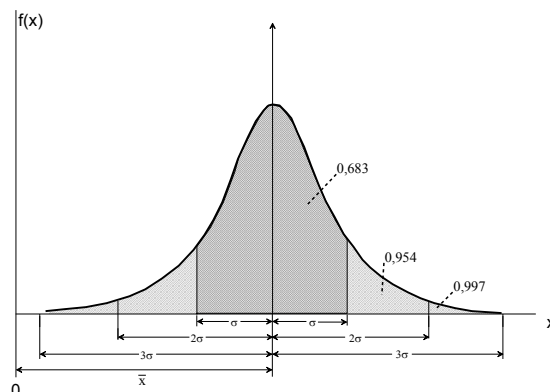
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\bar{x}-m)^2}{2\sigma^2}} \quad (27)$$

kde  $m$  je aritmetický priemer teoretického normálneho rozdelenia a  $\sigma$  je stredná kvadratická odchýlka (SKO) teoretického normálneho rozdelenia. Frekvenčná krivka normálneho rozdelenia má symetrický tvar. Maximum

$$f(x)_{\max} = \frac{1}{\sigma\sqrt{2\pi}} \quad (28)$$

je v bode  $x = m$ . Minimum je v bode  $x \rightarrow \pm\infty$ , t.j. asymptoticky sa blíži k vodorovnej osi.

**Obr.16.** Frekvenčná krivka normálneho rozdelenia. Na obrázku je vyznačená poloha aritmetického priemeru  $\bar{x}$  a intervaly, do ktorých s uvedenou pravdepodobnosťou padne ľubovoľný prvok súboru.



Pre aritmetický priemer a SKO normálneho rozdelenia platí

$$\begin{aligned} m &= \bar{x} \\ \sigma^2 &= s^2 \end{aligned} \quad (29)$$

Pre centrálny moment s-tého rádu normálneho rozdelenia platí

$$\mu_s = (s-1)\sigma^2\mu_{s-2} \quad (30)$$

a podľa tohto vzťahu totiž platí

$$\begin{aligned} \mu_1 &= 0 \\ \mu_2 &= \sigma^2 \\ \mu_3 &= 0 \\ \mu_4 &= 3\sigma^4 \end{aligned} \quad (31)$$

Z uvedeného je zrejmé, že všetky nepárne momenty sa rovnajú nule; párne momenty možno určiť podľa vzťahu

$$\mu_s = (s-1)!!\sigma^s \quad (32)$$

kde symbol  $(s-1)!!$  sa chápe ako násobok všetkých nepárnych čísel od 1 po  $s-1$ .

Keďže  $\mu_3 = 0$ , potom asymetria normálneho rozdelenia je rovná nule:

$$A = \frac{\mu_3}{\sigma^3} = 0 \quad (33)$$

a keďže pre štvrtý moment platí

$$\mu_4 = 3\sigma^4 \quad (34)$$

potom aj exces normálneho rozdelenia je rovný nule:

$$E = \frac{\mu_4}{\sigma^4} - 3 = \frac{3\sigma^4}{\sigma^4} - 3 = 0 \quad (35)$$

Logaritmicke normálne rozdelenie. Pri tomto rozdelení má normálne rozdelenie logaritmus náhodnej veličiny. Frekvenčná krivka LN-rozdelenia má tvar

$$f(\log x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} e^{-\frac{(\log \bar{x} - m)^2}{2\sigma^2}} \quad (36)$$

Aritmetický priemer LN-rozdelenia môžeme vypočítať zo vzťahu



$$\bar{x} = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} \quad (37)$$

LN-rozdelenie je asymetrické. V geologickej praxi sa často stretávame s nesymetrickým rozdelením. Nesymetrické rozdelenie môže vzniknúť vtedy, keď delenie parametra (vodorovná os) je z metodologického hľadiska použitá nevhodne. Takéto zdanlivo nesymetrické rozdelenie môže prejsť na symetrické, ak os parametra budeme deliť logaritmicky.

S logaritmicko-normálnym rozdelením sa stretávame v týchto prípadoch:

- a) pri práci s veľmi malými hodnotami (napr. nositeľmi magnetických vlastností hornín sú akcesorické minerály)
- b) pri štúdiu parameter, ktorého hodnoty zaberajú veľmi široký interval
- c) pri veľmi veľkej náhodnej chybe
- d) pri meraní času.

Ak je rozdelenie naozaj asymetrické, tak si asymetriu ponechá aj po zobrazení v logaritmickej stupnici.

Poissonovo rozdelenie. Majme náhodnú veličinu A, ktorá môže nadobúdať iba celé kladné hodnoty: 1; 2; 3; ... pričom postupnosť týchto hodnôt je teoreticky neohraničená. Hovoríme, že náhodná veličina A má Poissonovo rozdelenie, ak pravdepodobnosť P toho, že A nadobudne hodnotu x, je vyjadrená vzťahom

$$P_x = \frac{\lambda^x}{x!} e^{-\lambda} \quad (38)$$

kde  $\lambda$  je tzv. parameter Poissonovho rozdelenia (PR). Uvedený vzťah nám tiež udáva frekvenčnú funkciu PR. Pre aritmetický priemer a SKO PR platí

$$m = \sigma = \lambda \quad (39)$$

Táto vlastnosť PR sa v praxi používa na riešenie úlohy, či možno prijať hypotézu o tom, že náhodná veličina má PR. PR sa ako model (hypotéza) používa pri veľkom súbore ( $n > 30$ ) a malej pravdepodobnosti ( $p \leq 0,1$ ). PR je nesymetrické.

Fischerovo rozdelenie. V geologickej praxi sa niekedy používajú parametre vektorového typu (smery puklín, bridličnatosť, smery tektonických porúch, smery vektora prirodzenej remanentnej magnetizácie hornín a iné). Tieto vektory sa niekedy zobrazujú pomocou bodov na jednotkovej guľovej ploche [10]. V rovine sa táto plocha (a tiež body na nej) obyčajne zobrazuje stereografickou projekciou. Uvedené body sú dôležitým nositeľom informácie o vyšetřovanom parametri a teda, ako prvky určitého súboru, je potrebné ich štatisticky spracovať.

Keďže spôsob, ktorým sa smery vektorov zobrazujú, neumožňuje pre ich štatistické spracovanie použiť bežne používané štatistické modely a postupy, vypracoval R.A.Fischer [10] špeciálne rozdelenie a parametre, ktorými možno súbor smerov vektorov štatisticky popísať. Toto rozdelenie sa nazýva Fischerovo rozdelenie, resp. Fischerova štatistika.

Pre vysvetlenie funkcie Fischerovho rozdelenia použijeme súbor vektorov prirodzenej remanentnej magnetizácie vzoriek hornín  $\vec{J}_n$  :

Uvažujme ľubovoľnú magnetickú časticu – dipól - s magnetickým momentom  $M$  v hornine. Ak zanedbáme vzájomné pôsobenie medzi jednotlivými časticami, potom možno energiu častice ( $U$ ) v magnetickom poli zeme  $\vec{H}$  vyjadriť vzťahom

$$U = -MH\cos\psi \quad (40)$$

kde  $\psi$  je uhol medzi osou dipólu – častice a smerom magnetického poľa zeme  $\vec{H}$ .

Pravdepodobnosť  $dp$  toho, že energia  $U$  bude ležať v intervale  $\langle U; U + dU \rangle$  je

$$dp = C_1 e^{-\frac{U}{U_0}} dU = C_1 M H e^{-\frac{MH}{U_0} \cos\psi} \sin\psi d\psi = C e^{K \cos\psi} \sin\psi d\psi \quad (41)$$

kde  $U_0$  je energia náhodných fluktuácií (tepelných, hydrodynamických, ...);  $K$  je veličina, ktorá udáva, o koľko je maximálna energia častíc v zemskom magnetickom poli  $\vec{H}$  väčšia, ako energia fluktuácií. Táto veličina sa používa ako parameter hustoty smerov  $\vec{J}_n$ . Konštantu  $C$  možno určiť z podmienky, že ak funkciu  $dp$  budeme integrovať pre všetky možné hodnoty  $\psi$  a azimutálneho uhlu  $\varphi$ , musí byť tento integrál vždy rovný 1. Teda

$$\int_0^{2\pi} d\varphi \int_0^{\pi} C e^{K \cos\psi} \sin\psi d\psi = 1 \quad (42)$$

a odtiaľto je

$$C = \frac{K}{4\pi \text{sh}K} \quad (43)$$

a pre funkciu rozdelenia potom platí

$$p_0 = C e^{K \cos\psi} = \frac{K}{4\pi \text{sh}K} e^{K \cos\psi} \quad (44)$$

Zo vzťahu vidíme, že funkcia rozdelenia je priamo závislá na hodnote parametra  $K$ , ktorý je analogický parametru  $\sigma$  pre normálne rozdelenie.

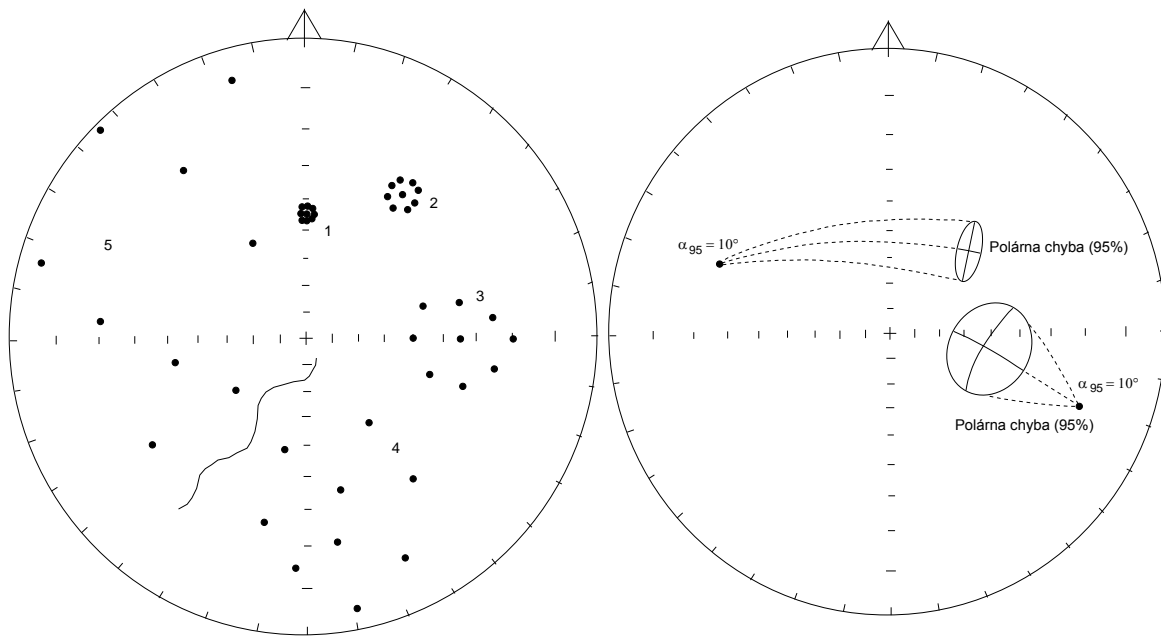
V bežnej praxi sa hodnota parametra  $K$  počíta podľa približného vzťahu

$$K = \frac{N-1}{N-R} \quad (45)$$

Kde  $N$  je počet vzoriek;  $R = \left[ \left( \sum_{i=1}^N \cos I_i \cos D_i \right)^2 + \left( \sum_{i=1}^N \cos I_i \sin D_i \right)^2 + \left( \sum_{i=1}^N \sin I_i \right)^2 \right]^{1/2}$  je „aritmetický priemer“ veľkosti smerov  $\bar{J}_n$ ;  $I_i$  je hodnota inklinácie veľkosti smerov  $\bar{J}_n$  i-tej vzorky;  $D_i$  je hodnota deklinácie veľkosti smerov  $\bar{J}_n$  i-tej vzorky. Hodnotu  $K$ , ktorú z uvedeného vzorca vypočítame, môžeme akceptovať len pri splnení podmienky  $K > 3$ . Polohu vektora  $R$  na jednotkovej guľovej ploche udávajú uhly:

$$I_R = \arcsin \frac{\sum_{i=1}^N \sin I_i}{R}$$

$$D_R = \arctg \frac{\sum_{i=1}^N \cos I_i \sin D_i}{\sum_{i=1}^N \cos I_i \cos D_i} \quad (46)$$



**Obr.17.** Príklad rôznej veľkosti rozptylu smerov (vľavo) a kružnice spoľahlivosti (vpravo)

Druhým parametrom Fischerovho rozdelenia je uhol  $\alpha$ . Je to uhol, ktorý je polovicou vrcholového uhlu kužľa, s vrcholom v strede jednotkovej guľovej plochy a základňou na guľovej ploche. Kružnica, ktorá tvorí obvod základne tohto kužľa sa nazýva **kružnica spoľahlivosti** a obsahuje s pravdepodobnosťou  $\pm p$  stredný smer súboru smerov  $\bar{J}_n$ .

Pre kosínus uhlu  $\alpha$  podľa Fischera platí vzťah

$$\cos\alpha = 1 - \frac{N-R}{R} \left[ \left( \frac{1}{p} \right)^{\frac{1}{N-1}} - 1 \right] \quad (47)$$

V paleomagnetickej praxi sa používa pravdepodobnosť  $1-p = 0,95$  a počíta sa podľa vzťahu

$$1 - \cos\alpha_{95} = \frac{N-R}{R} \left( 20^{\frac{1}{N-1}} - 1 \right) \quad (48)$$

$$\alpha_{95} = 81 \sqrt{\frac{N-R}{R} \left( 20^{\frac{1}{N-1}} - 1 \right)}$$

Kde  $N$  je počet vzoriek.

Hodnota  $\alpha$  pre pravdepodobnosť  $1-p = 1-1/e = 0,63$  sa používa ako SKO v určení stredného smeru  $\vec{J}_n$  a počíta sa podľa vzťahu

$$1 - \cos\alpha_{63} = \frac{N-R}{R} \left( e^{\frac{1}{N-1}} - 1 \right) \quad (49)$$

$$\alpha_{63} = 81 \sqrt{\frac{N-R}{R} \left( e^{\frac{1}{N-1}} - 1 \right)}$$

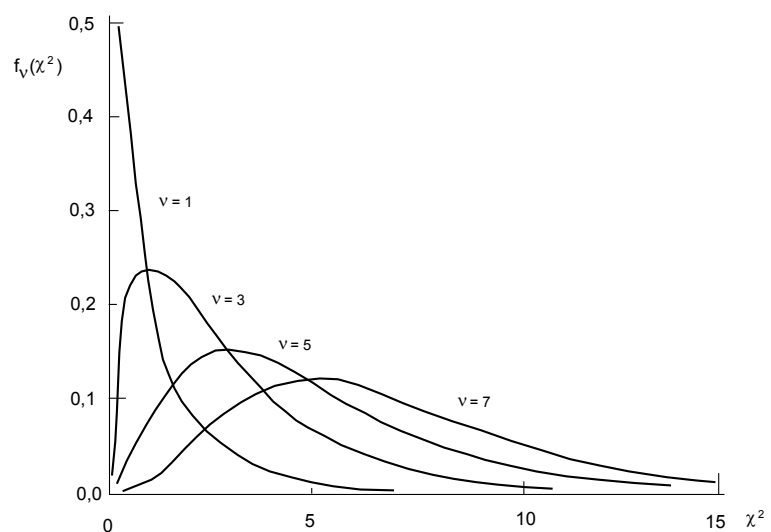
## 6. Intervaly spoľahlivosti a testy významnosti

Skôr než sa budeme zaoberať týmito charakteristikami, uvedieme si niektoré špeciálne rozdelenia, ktoré sa pri určovaní intervalov spoľahlivosti a pri testoch významnosti používajú.

Rozdelenie  $\chi^2$  (chí kvadrát).  
Dajme tomu, že zo základného súboru máme realizovaných niekoľko náhodných výberov. Jeden z týchto náhodných výberov má  $n$  prvkov ( $u_1, u_2, \dots, u_n$ ). Pre súčet kvadrátov (druhých mocnín) prvkov takéhoto súboru bol zavedený symbol

$$\chi^2 = \sum_{i=1}^n u_i^2 \quad (50)$$

Hodnota  $\chi^2$  môže pre každý výber nadobudnúť ľubovoľnú



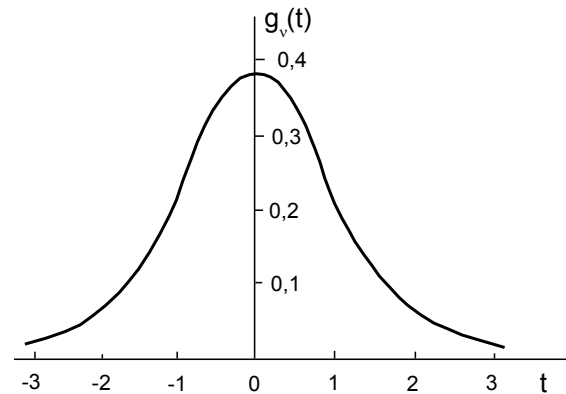
**Obr.18.** Frekvenčná funkcia rozdelenia  $\chi^2$  pre rôzne stupne voľnosti  $v$ .

hodnotu v intervale  $(0, \infty)$ . Je teda náhodnou veličinou, ktorá má svoje vlastné rozdelenie s frekvenčnou funkciou  $f_\nu(\chi^2)$  a distribučnou funkciou  $F_\nu(\chi^2)$ . Hodnoty týchto funkcií sú tabelované pre rôzne hodnoty  $\nu = n$ . Parameter  $\nu$  sa nazýva počet stupňov voľnosti a uvedená rovnosť platí len pre rozdelenie  $\chi^2$ . Hodnoty rozdelenia  $\chi^2$  pre určité pravdepodobnosti sa nazývajú kritické hodnoty rozdelenia  $\chi^2$ . Rozdelenie  $\chi^2$  je nesymetrické.

**Studentovo rozdelenie** (t-rozdelenie). Toto rozdelenie sa používa na hodnotenie odchýliek  $\bar{x} - m$  aritmetického priemeru náhodného výberu od aritmetického priemeru základného súboru. Tieto odchýlky sú definované ako náhodná veličina

$$t = \frac{\bar{x} - m}{s} \sqrt{n-1} \quad (51)$$

s t-rozdelením. t-rozdelenie je symetrické.



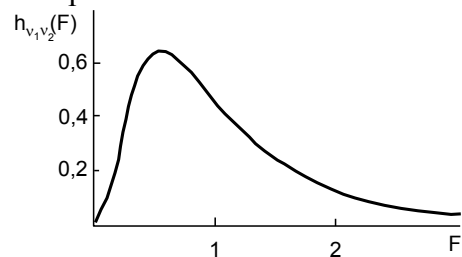
**Obr.19.** Frekvenčná krivka t-rozdelenia pre  $\nu = 5$ .

Rovnako ako je to pri rozdelení  $\chi^2$ , aj hodnoty t-rozdelenia sú tabelované. Kritické hodnoty t-rozdelenia udávajú pre danú pravdepodobnosť interval  $(-t, +t)$ , v ktorom bude náhodná veličina  $t$  ležať.

**F-rozdelenie** (Snedecorovo). Náhodná veličina  $F$  je definovaná pomerom

$$F = \frac{\chi_1^2}{\nu_1} \div \frac{\chi_2^2}{\nu_2} \quad (52)$$

Náhodná veličina  $F$  nadobúda iba kladné hodnoty a jej frekvenčná funkcia je nesymetrická. V praxi sa tento pomer volí tak, aby  $F \geq 1$ . S touto podmienkou pre rôzne



**Obr.20.** Frekvenčná krivka t-rozdelenia pre  $\nu_1 = 10$  a  $\nu_2 = 4$ .

stupne voľnosti sú hodnoty F-rozdelenia tabelované. Svoje kritické hodnoty veličina  $F$  neprekročí so zvolenou pravdepodobnosťou.

## 6.1 Intervaly spoľahlivosti

Pri štatistickom spracovaní určujeme štatistické charakteristiky náhodného výberu zo základného súboru. Otázkou je, s akou presnosťou vyjadrujú štatistické charakteristiky náhodného výberu, štatistické charakteristiky základného súboru. Túto otázku riešime pomocou intervalov spoľahlivosti. Interval spoľahlivosti nám hovorí, v akom intervale budú pre zvolenú pravdepodobnosť ležať štatistické charakteristiky základného súboru. Samozrejme, veľkosť intervalu spoľahlivosti vypočítame pomocou štatistických charakteristík náhodného výberu.

Pre **aritmetický priemer** základného súboru určíme interval spoľahlivosti podľa vzťahu

$$\bar{x} - t_p \frac{s}{\sqrt{n-1}} \leq m \leq \bar{x} + t_p \frac{s}{\sqrt{n-1}} \quad (53)$$

Tento vzťah platí pre súbory s malým rozsahom ( $n < 30$ ). Ak máme súbor s veľkým rozsahom, používame miesto  $t_p$  hodnotu  $u_p$  z normálneho rozdelenia.

Interval spoľahlivosti pre **medián** (nazýva sa tiež disperzný interval) je možné určiť pomocou grafu. Pre daný rozsah  $n$  vzostupne usporiadaného súboru sa určí z grafu poradové číslo dolnej hranice pomocou krivky  $x_d$  a poradové číslo hornej hranice pomocou krivky  $x_h$  na stupniciach  $x_d$ , resp.  $x_h$ . Zo súboru sa potom podľa poradových čísel vyberú hodnoty dolnej a hornej hranice intervalu spoľahlivosti.

Pre **strednú kvadratickú odchýlku** základného súboru určíme interval spoľahlivosti podľa vzťahu

$$\frac{ns^2}{\chi_{\frac{p}{2}}^2} \leq \sigma^2 \leq \frac{ns^2}{\chi_{1-\frac{p}{2}}^2} \quad (54)$$

index  $p$  pri hodnotách  $u_p$ ,  $t_p$  a  $\chi_p^2$  značí pravdepodobnosť, pre akú chceme interval spoľahlivosti počítať (napr. pre pravdepodobnosť 95% je  $p = 0,05$ ; pre 99% je  $p = 0,01$ ). Hodnoty  $u_p$ ,  $t_p$  a  $\chi_p^2$  možno nájsť v tabuľkách.

Intervaly spoľahlivosti sa tiež používajú na určenie náhodného výberu pre požadovanú presnosť

$$n = u_p^2 \frac{\hat{\sigma}^2}{\Delta^2} \quad (55)$$

kde  $u_p^2$  je tabuľková hodnota pravdepodobnosti s akou chceme meraný parameter určiť. Pre 95%-nú pravdepodobnosť je rovná 1,96, pre 99% je rovná 2,58;  $\hat{\sigma}^2$  je odhad SKO. Túto hodnotu volíme buď podľa predošlých meraní či z literatúry, alebo urobíme kontrolné (opakované) meranie. V tomto prípade hodnotu  $\hat{\sigma}^2$  vypočítame zo vzorca

$$\hat{\sigma}^2 = s^2 \frac{n}{n-1} \quad (56)$$

za  $\Delta$  dosadzujeme jednotku hodnoty požadovanej presnosti. Napr. 1 cm, 0,1 cm; 1 g a pod.

Pre potreby paleomagnetikkej praxe na určenie  $\Delta$  uviedol F. Janák [7] nasledovnú tabuľku (Tab.1 – stĺpec M v tabuľke udáva rozsah pre stredné hodnoty):

**Tab.1.** Hodnoty  $\Delta$  používané v paleomagnetickej praxi

M(10-6)	$\Delta$ (10-6)
0 – 10	10
10 – 100	30
100 – 500	100
500 – 1500	150
1500 – 5000	300
5000 - $\infty$	400

Iný spôsob na určenie potrebného počtu vzoriek uvádza N.P.Michajlová [8]

$$n = \left( \frac{1,96V\bar{x}}{100\lambda} \right) \quad (57)$$

Vzťah platí pre pravdepodobnosť  $p = 0,05$ ;  $V$  je variačný koeficient;  $\lambda$  je SKO aritmetického priemeru. Pre 20%-nú odchýlku  $\lambda$  upravila autorka uvedený vzťah na tvar

$$n_{20\%} \cong \frac{V^2}{100} \quad (58)$$

a na základe vlastných meraní potom určila nutný počet vzoriek: pre určenie susceptibility  $\kappa$  na

$$n_{20\%}(\kappa) = \frac{V^2}{100} = \frac{(30 \div 39)^2}{100} = 9 \div 15 \quad (59a)$$

a pre určenie magnetickej polarizácie  $J_n$  na

$$n_{20\%}(J_n) = \frac{V^2}{100} = \frac{(37 \div 49)^2}{100} = 14 \div 24 \quad (59b)$$

## 6.2 Testy významnosti

Testy významnosti sa používajú na overenie určitého predpokladu, napr.:

- či skúmaný výber pochádza zo základného súboru s určitým rozdelením
- či dva výbery pochádzajú z rovnakého základného súboru
- či je extrémna hodnota chybou merania, alebo patrí do základného súboru a pod.

Všeobecný postup pri použití testov významnosti má nasledovné kroky:

- voľba hladiny významnosti. Hladina významnosti nám udáva pravdepodobnosť, s ktorou bude splnené testovacie kritérium
- formulácia nulovej hypotézy. Testom napr. zisťujeme, či dva výbery pochádzajú z rovnakého základného súboru, t.j. že  $\bar{x}_1 = \bar{x}_2$ . Potom nulovú hypotézu formulujeme nasledovne:  $\bar{x}_1 - \bar{x}_2 = 0$ . Obyčajne testujeme rovnosť, resp. zhodu súborov.

Predpoklad tejto zhody sa matematicky formuluje rovnakou rovnicou ako bola posledne uvedená.

- 3) voľba testovacieho kritéria. Platnosť nulovej hypotézy sa hodnotí testovacím kritériom. Ako testovacie kritérium sa používajú niektoré teoretické rozdelenia, ktoré boli vyššie spomínané ( $\chi_p^2$ , t, F)
- 4) interpretácia výsledkov. Pri testovaní určíme výpočtom hodnotu zvoleného testovacieho kritéria. Túto hodnotu potom porovnáme z tabuľkovou hodnotou. V prípade, že vypočítaná hodnota je väčšia ako tabuľková kritická hodnota, nulovú hypotézu, na zvolenej hladine významnosti zamietneme a testovaný predpoklad považujeme za štatisticky významný, t.j. že nie je náhodný ale zákonitý. V prípade, že vypočítaná hodnota je menšia ako tabuľková, nulovú hypotézu na zvolenej hladine významnosti prijímame a testovaný predpoklad považujeme za štatisticky bezvýznamný, t.j. že je náhodný.

Test významnosti rozdielu medzi dvomi SKO. Podľa uvedenej schémy postupujeme nasledovne:

- 1) zvolíme hladinu významnosti  $p = (0,05, \text{ alebo } 0,01)$
- 2) formulujeme nulovú hypotézu  $\sigma_1^2 = \sigma_2^2$ , t.j. že SKO základných súborov sú rovnaké
- 3) ako testovacie kritérium volíme veličinu

$$F = \frac{\hat{\sigma}_1}{\hat{\sigma}_2} \quad (61)$$

- 4) z príslušnej tabuľky pre F (napr. [1], str.208) určíme kritické hodnoty pre  $p = 0,025$ , resp.  $p = 0,005$  a porovnáme ich s vypočítanou hodnotou. Ak je vypočítaná hodnota menšia, t.j. 95% resp. 99% z vypočítaných hodnôt F je menších ako  $F_{0,025}$ , resp.  $F_{0,005}$ , tak nulovú hypotézu  $\sigma_1^2 = \sigma_2^2$  prijmem. V opačnom prípade nulovú hypotézu zamietneme, t.j.  $\sigma_1^2 \neq \sigma_2^2$ .

Tento test používame pre súbory, ktoré majú aspoň približne normálne rozdelenie.

t-test, pre prípad, že  $\sigma_1^2 = \sigma_2^2$  (pre testovanie zhody dvoch aritmetických priemerov). Postup aplikácie testu je nasledovný:

- 1) zvolíme hladinu významnosti  $p = (0,05, \text{ alebo } 0,01)$
- 2) pre oba skúmané výbery vypočítame  $\bar{x}_1, s_1^2, \bar{x}_2, s_2^2$
- 3) F-testom zistíme, či netreba zamietnuť nulovú hypotézu  $\sigma_1^2 = \sigma_2^2$
- 4) podľa vzťahu

$$t = A_t \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \quad (62)$$

vypočítame hodnotu testovacieho kritéria t a stanovíme počet stupňov voľnosti  $v = n_1 + n_2 - 2$

- 5) z príslušnej tabuľky vyhľadáme pre dané v hodnotu  $t_p$  a porovnáme ju s vypočítanou hodnotou testovacieho kritéria t.



- 6) Ak  $t > t_p$ , potom zamietame nulovú hypotézu  $m_1 = m_2$  a tvrdíme, že rozdiel aritmetických priemerov je štatisticky významný na hladine významnosti  $p$ . V opačnom prípade nulovú hypotézu prijímame a rozdiel priemerov považujeme za nevýznamný

Test sa môže použiť pre nepárové hodnoty za predpokladu normálneho rozdelenia oboch základných súborov.

t-test, pre prípad, že  $\sigma_1^2 \neq \sigma_2^2$  (pre testovanie zhody dvoch aritmetických priemerov). Postup aplikácie testu je nasledovný:

- 1) štatistickým charakteristikám oboch súborov pridáme indexy tak, aby  $\sigma_1^2 > \sigma_2^2$
- 2) zvolíme hladinu významnosti  $p = (0,05, \text{ alebo } 0,01)$
- 3) F-testom dokážeme, že zamietame nulovú hypotézu  $\sigma_1^2 = \sigma_2^2$ , to znamená, že prijímame  $\sigma_1^2 \neq \sigma_2^2$
- 4) Vypočítame hodnotu testovacieho kritéria  $t$  podľa vzťahu

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}} \quad (63)$$

- 5) z príslušnej tabuľky (napr. [1], str.207) vyhľadáme hodnoty  $t'_p$  a  $t''_p$  pre  $v_1 = n_1 - 1$  a  $v_2 = n_2 - 1$  stupňov voľnosti
- 6) vypočítame kritickú hodnotu  $t_p^*$  podľa vzťahu

$$t_p^* = \frac{t'_p \frac{s_1^2}{n_1 - 1} + t''_p \frac{s_2^2}{n_2 - 1}}{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}} \quad (64)$$

- 7) nulovú hypotézu  $m_1 = m_2$  zamietame, keď  $t > t_p^*$ ; v tom prípade tvrdíme, že rozdiel aritmetických priemerov je na hladine významnosti  $p$  významný.

Test zhody pre jeden výber (Pearsonov test). Pomocou testov zhody sa určuje zhoda dvoch rozdelení, t.j. či výber pochádza z určitého základného súboru alebo nie, či dva výbery pochádzajú z rovnakého základného súboru, či určitý teoretický základný súbor môže byť modelom pre študovaný výber a pod.

Pri teste zhody pre jeden výber, rozdelíme si testovaný výber na triedy. Početnosti v jednotlivých triedach označíme ako experimentálne početnosti  $n_{e,j}$ . Tieto experimentálne početnosti budeme porovnávať s určitým rozdelením, pre ktoré určíme očakávané (modelové) početnosti  $n_{0,j}$ . Postup je ďalej nasledovný:

- 1) určíme hladinu významnosti
- 2) formulujeme nulovú hypotézu  $n_{e,j} = n_{0,j}$

- 3) určíme testovacie kritérium (testovacích kritérií existuje niekoľko – podľa toho, aké špeciálne rozdelenie použijeme) napr.:

$$\chi^2 = \sum_{j=1}^k \frac{n_{e,j} - n_{0,j}}{n_{0,j}} \quad (65)$$

kde k je počet tried; toto testovacie kritérium má  $v = k - 1$  stupňov voľnosti.

- 4) vypočítame očakávané početnosti (napr. [1], str.94) a hodnotu testovacieho kritéria. Porovnáme vypočítané hodnoty s tabuľkovými a prijmem, alebo zamietneme nulovú hypotézu.

Test zhody pre dva nezávislé výbery (Kolmogorovov – Smirnovov test). Týmto testom určíme, či dva testované výbery pochádzajú z rovnakého základného súboru, alebo nie. Tento test sa používa:

- pre malé výbery, ak  $n_1 = n_2 \leq 40$
- pre veľké výbery, ak  $n_1 > 40 < n_2$

Postup v prípade a):

- určíme nulovú hypotézu:  $N_{1,j} = N_{2,j}$  ( $N_{1,j}$  a  $N_{2,j}$  sú kumulatívne početnosti)
- zvolíme hladinu významnosti  $p$
- zvolíme testovacie kritérium –  $D_2 = \text{maximum } |N_{1,j} - N_{2,j}|$  (66)
- vykonáme výpočet kumulatívnych početností v oboch výberoch, vypočítame rozdiely príslušných dvojíc. Najväčší rozdiel je testovacím kritériom. Vypočítanú hodnotu porovnáme s tabuľkovou (napr. [1], str.232). Ak je vypočítaná hodnota väčšia, alebo rovná tabuľkovej, nulovú hypotézu zamietame.

Postup v prípade b):

Tento je podobný ako v prípade a) s tým rozdielom, že sa používajú relatívne početnosti a relatívne kumulatívne početnosti. Relatívnu početnosť vypočítame zo vzťahu:

$$f_i = \frac{n_i}{n} \quad (67)$$

Kde  $n_i$  je početnosť v  $i$ -tej triede;  $n$  je rozsah súboru; relatívne kumulatívne početnosti počítame z relatívnych početností. Ďalší rozdiel je v tom, že tabuľkovú hodnotu si musíme vypočítať, pretože nie je tabelovaná. Pre  $p = 0,05$  je

$$D_{2;0,05} = 1,36 \sqrt{\frac{n_1 + n_2}{n_1 \times n_2}} \quad (68)$$

Pre  $p = 0,01$  je

$$D_{2;0,01} = 1,63 \sqrt{\frac{n_1 + n_2}{n_1 \times n_2}} \quad (69)$$

Nulová hypotéza sa zamieta vtedy, ak  $D_2 \geq D_{2;p}$ .

Test extrémnych odchýliek. Testy podobného typu používame vtedy, keď vo výberovom súbore hodnôt sa vyskytuje extrémne veľká či malá hodnota, ktorá vzbudzuje nedôveru. Aby sme zistili, či patrí do základného súboru alebo nie (patrí do iného základného súboru, alebo je chybou merania), vykonáme nasledovné testovanie:

- 1) zvolíme hladinu významnosti  $p = (0,05, \text{ alebo } 0,01)$
- 2) formulujeme nulovú hypotézu:  $T_n \leq T_{\text{tab.}}$ ; resp.  $T_1 = T_{\text{tab.}}$ .
- 3) Zvolíme testovacie kritérium

$$T_n = \frac{x_n - \bar{x}}{s} \qquad T_1 = \frac{\bar{x} - x_1}{s} \qquad (70)$$

kde  $T_n$  je testovacie kritérium pre extrémne veľkú hodnotu;  $x_n$  je najväčšia hodnota podľa veľkosti usporiadaného súboru;  $T_1$  je testovacie kritérium pre extrémne malú hodnotu;  $x_1$  je najmenšia hodnota podľa veľkosti usporiadaného súboru

- 4) vykonáme výpočet testovacieho kritéria. Ak je vypočítaná hodnota väčšia ako tabuľková (napr. [1], str.233), nulovú hypotézu zamietame.

Test náhodnosti výberu. Na zistenie predpokladu, že výber z nejakého základného súboru je náhodný, sa niekedy používa test, založený na kritických hodnotách pomeru strednej kvadratickej postupnej diferencie a výberovej SKO.

Postup testovania je nasledovný: predpokladáme, že testovaný výber je zo základného súboru s normálnym rozdelením. Získané prvky výberu nezaraďujeme podľa veľkosti, ale ponechávame náhodný sled hodnôt. Vypočítame SKO výberu podľa vzťahu

$$s'^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \qquad (71)$$

Ďalej vypočítame strednú kvadratickú postupnú diferenciu podľa vzťahu

$$\Delta^2 = \frac{1}{n-1} \sum_{i=1}^N (x_{i+1} - x_i)^2 \qquad (72)$$

Pri náhodnom súbore by mal byť pomer

$$\frac{\Delta^2}{s'^2} = 1,0 \qquad (73)$$

Tento stav však nemôže nastať a preto sa povoľuje určitá tolerancia. To znamená, že pre zvolenú pravdepodobnosť  $1-p$  (hladina významnosti) a zvolený počet stupňov voľnosti ( $v = n$ ) existujú kritické hodnoty, ktoré pomer (73) nesmie prekročiť, aby sa zachoval predpoklad náhodnosti výberu. Číže ak nami vypočítaná hodnota

$$K_{\text{vyp}} = \frac{\Delta^2}{s'^2} \quad (74)$$

je väčšia ako tabuľková –  $K_K$  (možno ju určiť z grafu na obr. 26 [1]) hovoríme, že výber je náhodný. V opačnom prípade, t.j. ak

$$K_{\text{vyp}} < K_K \quad (75)$$

Hovoríme, že výber nie je náhodný na zvolenej hladine významnosti.

### 6.3 Vyrovnávanie štatistických dát

Do štatistickej analýzy určitého štatistického súboru patrí určenie modelového rozdelenia pre tento študovaný súbor.

Študovaný súbor zobrazíme vo forme histogramu. Teoretické rozdelenie pomocou frekvenčnej krivky. Určenie modelového rozdelenia závisí v preložení histogramu vhodnou teoretickou krivkou. Tomuto hovoríme vyrovnávanie štatistických dát.

Po vyrovnaní štatistických dát sa ešte musíme presvedčiť, či sme správne zvolili teoretické rozdelenie. Túto kontrolu robíme pomocou testov zhody.

#### Vyrovnávanie pomocou normálneho rozdelenia.

- 1) v prípade väčších súborov robíme vyrovnávanie experimentálnych hodnôt pomocou normálneho rozdelenia pomocou vzťahu

$$\varphi(x) = \frac{nh}{s\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2s^2}} \quad (76)$$

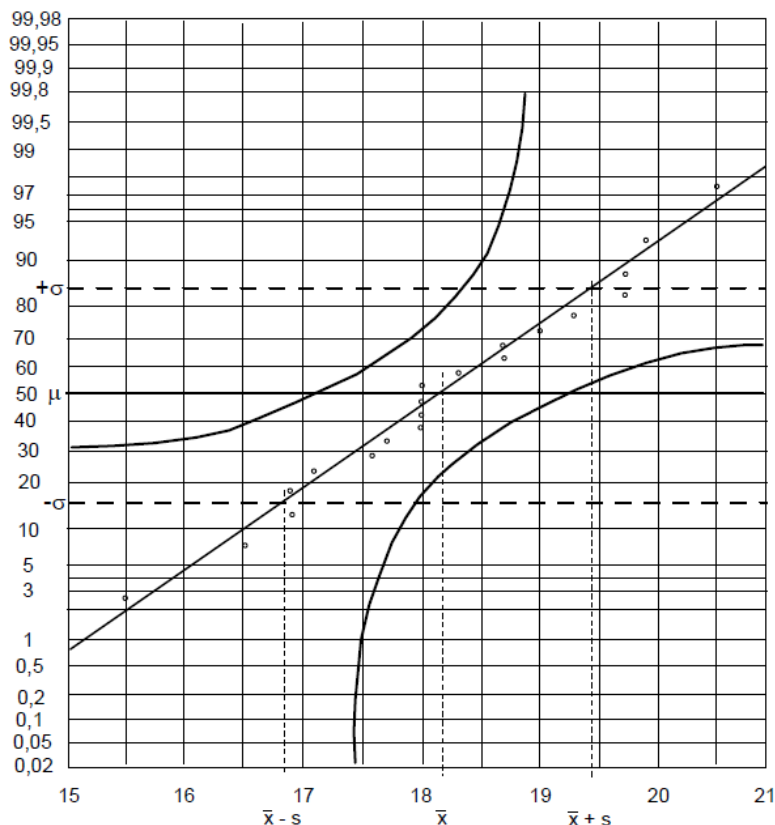
kde  $n$  je počet prvkov súboru;  $h$  je veľkosť triedneho intervalu;  $s$  je SKO;  $\bar{x}$  je aritmetický priemer. Vhodnosť modelu testujeme niektorým z testov zhody (napr. [1], str.126)

- 2) v prípade, že máme malý súbor, ktorý sme nerozdelili do tried, na vyrovnávanie používame pravdepodobnostný papier. Táto metóda má však iba približný, orientačný charakter. Pri použití postupujeme zhruba nasledovne: zostavíme si tabuľku, v ktorej budú podľa veľkosti usporiadané hodnoty súboru a im odpovedajúce relatívne kumulatívne početnosti. Tieto vypočítame podľa vzťahu:  
pre  $i$ -tu hodnotu platí

$$F_{(i)} = \left(i - \frac{1}{2}\right) \frac{100}{n} (\%) \quad (77)$$

Hodnoty tabuľky vo forme bodov vynesieme na pravdepodobnostný papier. Na horizontálnu os nanášame hodnoty súboru, na vertikálnu os im odpovedajúce relatívne kumulatívne početnosti. Potom body „od oka“ preložíme priamkou. Táto priamka predstavuje modelové rozdelenie pre bodmi zobrazený súbor. Zhodu rozdelení (experimentálneho a teoretického) testujeme (Kolmogorovov – Smirnovov test).

Tabuľkovú hodnotu pre daný počet prvkov a zvolenú hladinu významnosti, vynásobenú 100, vyznačíme na obe strany priamky v mierke zvislej osi. Ak len jeden bod bude ležať mimo vyznačených hraníc, nulovú hypotézu (že experimentálny súbor má normálne rozdelenie) zamietneme.



**Obr.21.** Pravdepodobnostný papier s vyznačeným testovaným súborom (body)

Vyrovnanie pomocou Poissonovho rozdelenia. Výpočet modelového rozdelenia robíme podľa vzťahu

$$\varphi(x_j) = nf(x_j) \quad (78)$$

kde  $n$  je rozsah experimentálneho súboru a  $f(x_j)$  je hodnota Poissonovho rozdelenia. Určíme ju nasledovným spôsobom: vypočítame aritmetický priemer súboru -  $\bar{x}$ . Táto hodnota je aritmetickým priemerom Poissonovho rozdelenia

$$\bar{x} = \lambda \quad (79)$$

potom pre hodnoty  $\lambda$  a jednotlivé hodnoty prvkov súboru -  $x_j$  z tabuľky (napr. [1], str.198), nájdeme hodnoty  $f(x_j)$ . Zhodu experimentálneho a modelového súboru zistíme jedným z testov zhody (napr. [1], str.137).

## 7 Disperzná analýza

Výsledky sledovaného javu často závisia od mnohých nezávislých faktorov, meniacich svoju hodnotu počas pokusu. Kvantitatívne hodnotenie vplyvu jednotlivých faktorov na študovaný jav pri určitých podmienkach môže byť vykonané pomocou disperznej analýzy.

### 7.1 Jednofaktorová disperzná analýza

Pre názornosť si uvedieme jednoduchý príklad: treba zhodnotiť hypotézu, že faktor B vplýva na výsledky meraní. Predpokladajme, že faktor B môže nadobudnúť  $r$  rôznych hodnôt. Pre každú z týchto hodnôt ( $B_1, B_2, \dots, B_r$ ) realizujeme výber, ktorý má ( $n_1, n_2, \dots, n_r$ ) prvkov a jeho aritmetický priemer je ( $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_r$ ). Tieto hodnoty si zostavíme do tabuľky:

**Tab.2.** Tabuľka pre vykonanie disperznej analýzy

Hodnoty faktora B	Jednotlivé prvky	Počet prvkov	Aritm. Priemer
$B_1$	$x_{11}, x_{12}, x_{13}, \dots, x_{1n_1}$	$n_1$	$\bar{x}_1$
$B_2$	$x_{21}, x_{22}, x_{23}, \dots, x_{2n_2}$	$n_2$	$\bar{x}_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$B_r$	$x_{r1}, x_{r2}, x_{r3}, \dots, x_{rn_r}$	$n_r$	$\bar{x}_r$

Celkový počet prvkov je

$$n = \sum_{i=1}^r n_i \quad (80)$$

Aritmetický priemer celého súboru je

$$\bar{x} = \frac{1}{n} \sum_{i=1}^r n_i \bar{x}_i \quad (81)$$

Teraz vyslovíme nulovú hypotézu:

predpokladajme, že faktor B nevplýva na výsledky meraní, t.j. že ak  $\bar{x}_0 = a$ , potom

$$a_1 = a_2 = \dots = a_r = a \quad (82)$$

Aby sme mohli určiť interval spoľahlivosti tejto hypotézy, musíme určiť SKO aritmetických priemerov jednotlivých hodnôt B. Výpočtom ([2], str.80) zistíme, že táto odchýlka sa skladá z dvoch hodnôt:

$$\sigma^2 = \frac{1}{n-1} \sum (x_i - \bar{x}_i)^2 \quad (83)$$

a

$$\sigma_b^2 = n(\bar{x} - a)^2 \quad (84)$$

V prvom prípade  $\sigma^2$  nezávisí od **a** a aritmetický priemer jednotlivých hodnôt B nebude rovný **a**. V druhom prípade bude mať  $\sigma_b^2$  zmysel iba v tom prípade, ak aritmetický priemer jednotlivých hodnôt B bude rovný **a**. Ak aritmetický priemer jednotlivých hodnôt B sa líši od aritmetického priemeru celého súboru hodnôt B, potom pomer

$$F = \frac{\sigma_b^2}{\sigma^2} \quad (85)$$

ktorý má F-rozdelenie, bude väčší ako 1. Veličina F teda môže slúžiť ako kritérium nulovej hypotézy.

Prakticky sa obe hodnoty SKO počítajú podľa vzťahov

$$\sigma^2 = \frac{1}{n-r} \sum_{i=1}^R \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \quad (86)$$

Tento vzorec predstavuje neskrútené hodnotenie SKO, ktorá vzniká kolísaním prvkov  $x_{ij}$  vo vnútri jednotlivých výberov. Druhá hodnota

$$\sigma_b^2 = \frac{1}{r-1} \sum_{i=1}^r n_i (\bar{x}_i - \bar{x})^2 \quad (87)$$

Vyjadruje kolísanie aritmetického priemeru jednotlivých hodnôt B okolo aritmetického priemeru celého súboru B.

Správnosť hypotézy kontrolujeme pomocou F-rozdelenia. Vypočítame pomer

$$F = \frac{\sigma_b^2}{\sigma^2} \quad (88)$$

Hodnotu F-rozdelenia nájdeme v tabuľkách (napr. [1], str.208) pre  $r - 1$  a  $n - r$  stupňov voľnosti. Táto tabuľková hodnota nám ohraničuje oblasť možných hodnôt vypočítaného pomeru F. Ak vypočítaná hodnota F je väčšia ako tabuľková znamená to, že faktor B vyplýva na výsledky merania.

## 7.2 Dvojfaktorová disperzná analýza

Tak ako existuje jednofaktorová disperzná analýza, existuje samozrejme aj dvojfaktorová disperzná analýza. Používa sa na určenie vplyvu dvoch faktorov (A a B) na študovaný jav.

Na tomto mieste sa dvojfaktorovou disperznou analýzou zaoberať nebudeme. Prípadných záujemcov odkazujeme na prácu [2], str.84 a [14].

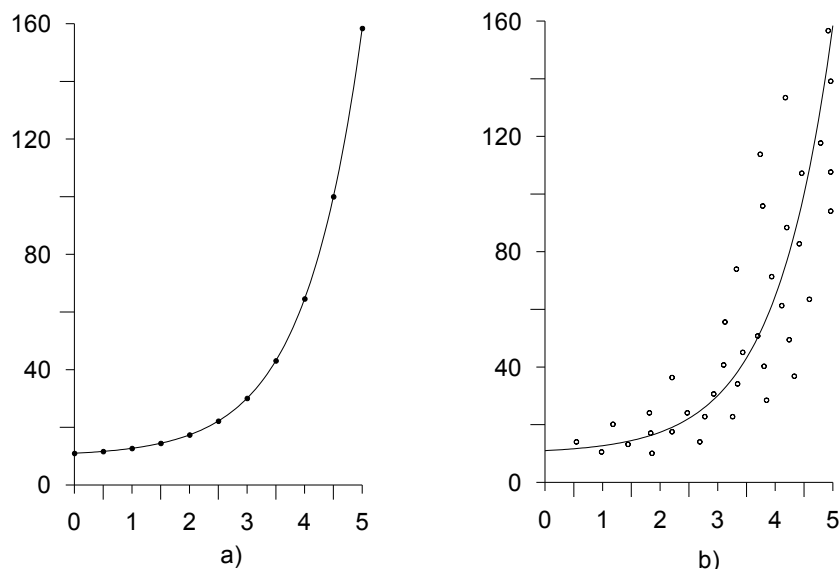
## 8. Závislosť náhodných veličín

Zatiaľ sme sa zaoberali štúdiom jednorozmerných štatistických súborov, t.j. takých súborov, kde na každej štatistickej jednotke (na každom prvku) bol sledovaný iba jeden štatistický znak. V praxi však zvyčajne pracujeme s viacrozmernými štatistickými súbormi, pri ktorých na každej štatistickej jednotke zistíme väčší počet znakov.

Keby sme každý znak študovali izolovane od ostatných, pripravili by sme sa o dôležitú informáciu, vyplývajúcu z ich vzájomného vzťahu. Žiadny jav v prírode nevzniká ani neprebíha ľubovoľne, ale je vo vzťahu k iným javom a nemôže byť správne pochopený, ak je z týchto vzťahov a súvislostí vytrhnutý. Medzi javmi teda objektívne existujú príčinné vzťahy a závislosti, ktorých štúdium patrí k základným úlohám každého vedeckého skúmania. Príčinou závislostí pritom rozumieme takú súvislosť javov, kde jeden z nich – ako **príčina** – vyvoláva nutne druhý – ako **následok**, prípadne, kedy sa javy podmieňujú vzájomne.

Závislosť medzi náhodnými veličinami môže byť v zásade dvojakého typu: **funkčná** a **stochastická** (štatistická, pravdepodobnostná). **Funkčná závislosť** sa prejavuje jednoznačne a presne pri každom pozorovaní. Je charakteristická tým, že určitej hodnote nezávisle premennej odpovedá jediná hodnota závisle premennej.

Pri **stochastickom vzťahu** však jednej hodnote nezávisle premennej odpovedá viacero hodnôt závisle premennej a ich možné číselné hodnoty môžeme iba s určitou pravdepodobnosťou predpovedať. Pre **štatistickú závislosť** je teda príznačné, že zmenám hodnôt jednej veličiny, odpovedá zmena priemeru rozloženia početností hodnôt druhej veličiny. Najlepšie si to ukážeme na obr.22a a 22b. Na obr.22a je zobrazená funkčná závislosť a na obr.22b je zobrazená štatistická závislosť. Štatistickú závislosť tiež nazývame **korelačná závislosť**. Stochastický vzťah vzniká medzi náhodnými veličinami obyčajne vtedy, keď ich ovplyvňujú všeobecné náhodné faktory spolu s inými náhodnými faktormi, rôznymi pre každú náhodnú veličinu.



**Obr.22a** Grafické vyjadrenie funkčného vzťahu  $y = a + e^x$

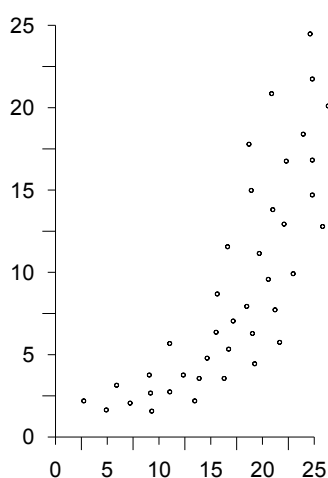
**Obr.22b** Grafické vyjadrenie štatistickej závislosti



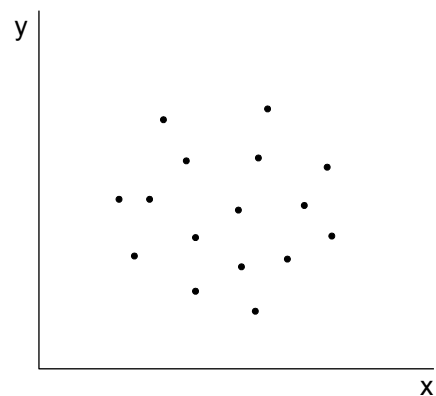
V praxi sa často stretávame s prípadmi, keď sa na jednom prvku súčasne meria viac veličín, ktoré môžu byť medzi sebou závislé. Našou úlohou je potom tieto závislosti hľadať. Štatistickými závislosťami sa zaoberá časť matematickej štatistiky, nazývaná **korelačná analýza**. Na funkčnú závislosť sa potom môžeme pozerieť ako na limitný prípad štatistickej závislosti, ku ktorej by sme sa mohli dopracovať elimináciou všetkých náhodných vplyvov.

Korelačná analýza má dve základné úlohy:

- 1) popísať kvantitatívny priebeh skúmaného vzťahu a použiť korelačnú závislosť pri odhadoch zmien výslednej veličiny. To je **regresná úloha** korelačnej analýzy. Jej výsledky sa používajú pre prognózach.
- 2) Druhou základnou úlohou korelačnej analýzy je skúmanie **tesnosti korelačného vzťahu**. Táto úloha sa nazýva **korelačná**.



**Obr.23.** Bodový graf



**Obr.24.** Nekorelovaný vzťah dvoch veličín

### 8.1 Regresná úloha korelačnej analýzy

Korelačná závislosť sa najlepšie zobrazuje bodovým grafom (Obr.23).

Iným prostriedkom na poznanie korelačnej závislosti je tzv. korelačná tabuľka (Tab.3).

V hlavičkách korelačnej tabuľky sú uvedené intervaly oboch premenných a v poli sú uvedené početnosti prislúchajúce patričnej dvojici intervalov.

Teraz si povieme niečo o metódach riešenia regresnej úlohy. Pôjde tu o to aby sme vyjadrili priebeh korelačnej závislosti t.j. zmeny závisle premennej, spojené so zmenami nezávisle premennej. Tento vzťah nazývame **regresia**.

Experimentálne získavame materiál o regresnej závislosti dvoch premenných. Tieto experimentálne údaje však dostatočne nevystihujú skúmanú závislosť, pretože na ich hodnoty pôsobia náhodné vplyvy. Regresiu preto vyjadrujeme **regresnou funkciou**. Je to matematická funkcia ktorá vyjadruje hlavnú tendenciu závislosti. Je nutné si na tomto mieste uvedomiť, že pravdepodobnosť, s akou bude regresná funkcia vyjadrovať vzťah dvoch premenných, je priamo závislá na rozsahu súboru.

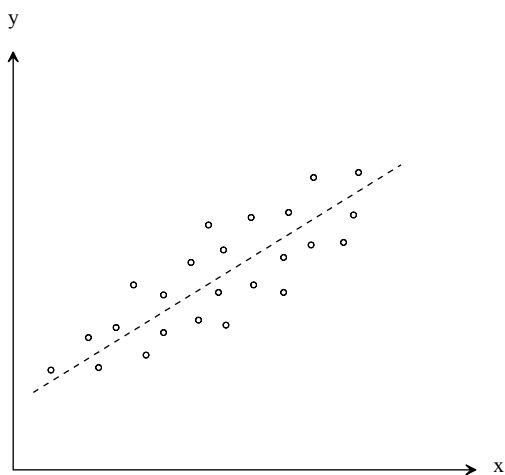
**Tab.3. Korelačná tabuľka**

x \ y	75-99	100-124	125-149	150-174	175-199	200-224	225-249	250-274	275-299	n <sub>j</sub>
750-999	1	-	-	1	-	-	-	-	-	2
1000-1249	-	1	-	1	3	-	-	-	-	5
1250-1499	-	1	1	5	5	1	-	-	-	13
1500-1749	-	1	2	4	6	1	-	-	-	14
1750-1999	-	1	2	4	9	4	-	-	-	20
2000-2249	-	-	2	2	5	5	3	1	-	18
2250-2499	-	-	-	2	4	4	6	-	-	16
2500-2749	-	-	-	-	2	3	6	-	-	11
2750-2999	-	-	-	-	-	2	2	1	1	6
3000-3249	-	-	-	-	-	1	2	-	-	3
3250-3499	-	-	-	-	-	-	-	1	1	2
n <sub>i</sub>	1	4	7	19	34	21	19	3	2	n=110

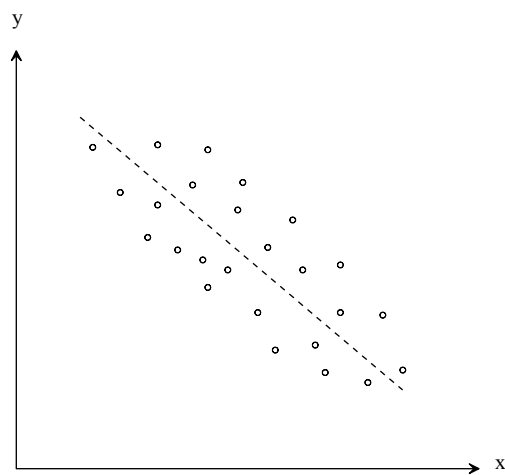
Lineárna regresia. Najjednoduchšou formou regresie je priamková alebo lineárna regresia. Regresnú priamku vyjadrujeme rovnicou

$$Y = a_{yx} + b_{yx}X \quad (89)$$

veľké písmená (Y a X) používame preto, aby sme odlíšili teoretické hodnoty od praktických, ktoré označujeme malým písmenom y;  $a_{yx}$  a  $b_{yx}$  sú konštanty priamky. Indexy vyjadrujú, že ide o závislosť y na x. Zvláštny význam má koeficient  $b_{yx}$ , nazývaný tiež regresný koeficient. Ak je kladný, znamená to, že ide o regresiu kladnú – obr.25a, ak je záporný znamená to, že ide o regresiu zápornú – obr.25b.



a) kladná regresia



b) záporná regresia

**Obr.25. Kladná a záporná regresia**

Ak  $b_{yx} = 0$ , hovoríme o korelačnej nezávislosti oboch premenných. Koeficienty  $a_{yx}$  a  $b_{yx}$  určíme podľa nasledovných vzťahov

$$b_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} \quad (90)$$

$$a_{yx} = \bar{y} - b_{yx} \bar{x} \quad (91)$$

kde  $\overline{xy}$  je aritmetický priemer zo súboru súčinov ( $x_i \times y_i$ );  $\bar{x}$  je aritmetický priemer zo súboru hodnôt  $x$ ;  $\bar{y}$  je aritmetický priemer zo súboru hodnôt  $y$ .

Dôležitým ukazovateľom korelačnej závislosti je činiteľ z rovnice (90). Nazývame ho **kovariácia** a označujeme symbolom **cov xy**

$$\text{cov xy} = \overline{xy} - \bar{x} \cdot \bar{y} \quad (92)$$

Ak je kovariácia kladná znamená to, že obe premenné (závisle i nezávisle) sa zväčšujú. V prípade že je záporná znamená to, že so zväčšovaním jednej premennej druhá premenná sa znižuje.

Niekedy sa stretávame s takou korelačnou závislosťou, pri ktorej sú obidve premenné závisle premenné na nejakej inej veličine. Medzi prvými dvomi potom existuje tzv. vzájomná korelačná závislosť. V takom prípade existujú dve regresné priamky a dve dvojice regresných ukazovateľov. Hovoríme potom o **združených regresných priamkach**. Teda okrem regresnej priamky  $Y$ , existuje aj s ňou združená regresná priamka

$$x = a_{xy} + b_{xy}y \quad (93)$$

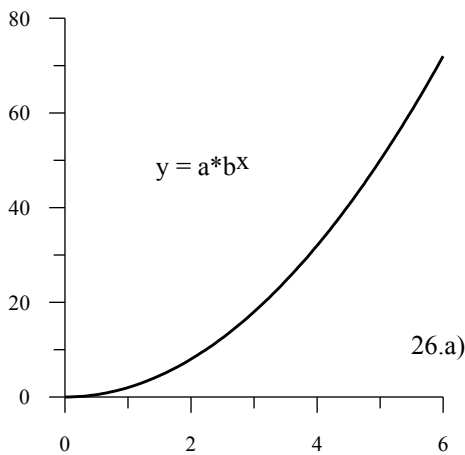
S parametrami

$$b_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{y^2} - \bar{y}^2} \quad (94)$$

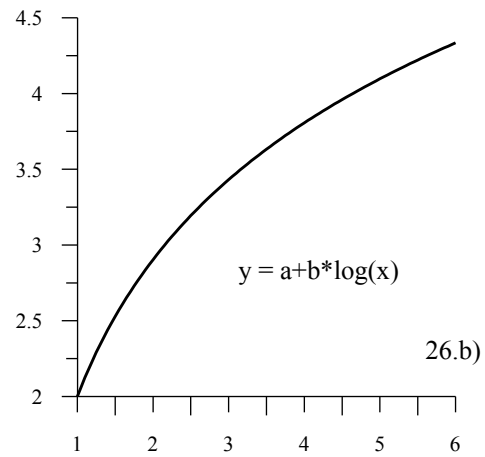
$$a_{xy} = \bar{x} - b_{xy} \bar{y} \quad (95)$$

Nelineárna regresia. V podstate sa stretávame s týmito typmi nelineárnej regresie [4]:

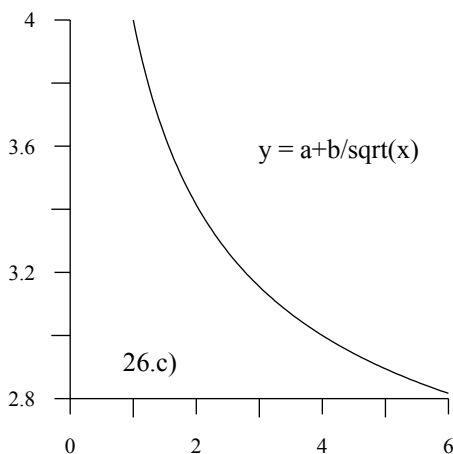
- s rastom hodnôt nezávisle premennej rastú aj hodnoty závisle premennej a to tak, že prírastky sa stále zväčšujú (obr.26. a))
- rovnako ako v prípade a), ale prírastky sa stále znižujú (obr.26. b))
- s rastom hodnôt nezávisle premennej klesajú hodnoty závisle premennej, pričom tento pokles sa stále spomaľuje (obr.26. c))
- rovnako ako v prípade c), ale pokles sa stále zrýchľuje (obr.26. d))



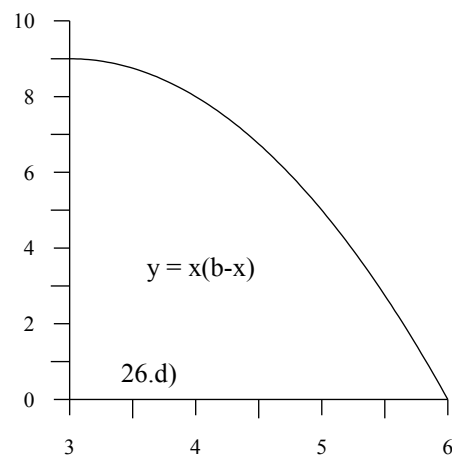
**Obr.26. a) Nelineárna regresia**



**Obr.26. b) Nelineárna regresia**



**Obr.26. c) Nelineárna regresia**



**Obr.26. d) Nelineárna regresia**

Funkcií, ktoré sú v určitom intervale schopné vyjadriť niektoré z uvedených typov nelineárnej regresie je obyčajne viac. Z praktických dôvodov sa volia také, ktorých konštanty sa dajú z experimentálnych údajov ľahko určiť. Voľba regresnej funkcie je obmedzená niekoľkými faktormi; jedným z najdôležitejších je rozsah súboru. Počet parametrov regresnej funkcie musí byť mnohonásobne menší, ako počet párov hodnôt korelačných veličín. Rovnako musí byť počet parametrov regresnej funkcie menší ako počet skupín (intervalov, tried).

## 8.2 Korelačná úloha

Skôr než uvedieme formy hodnotenia tesnosti korelačnej závislosti, je potrebné uviesť podmienky, za akých možno tento korelačný vzťah skúmať. Totiž, na zmeny jednej závislej premennej veličiny, môžu vplývať viaceré nezávislé premenné veličiny, ktoré môžu byť buď navzájom korelované, alebo nie. Prítom za vzájomne nekorelované veličiny považujeme tie, ktoré majú nulovú kovariáciu. Teda:

- v prípade, že nezávislé premenné sú navzájom nekorelované, potom rozptyl hodnôt závislej premennej okolo regresnej funkcie  $Y$  je spôsobený len náhodným pôsobením tých nezávislých veličín pre ktoré korelačnú závislosť nezisťujeme.
- V prípade, že nezávislé premenné sú navzájom korelované potom jednak regresná funkcia a tiež aj rozptyl závislej premennej okolo regresnej funkcie  $Y$  je ovplyvnený týmito korelačným vzťahom. Obe charakteristiky sú skreslené.

Hodnoty závisle premennej vykazujú určitú premenlivosť (variabilitu), ktorú vyvolávajú ako znaky a tiež zmeny ostatných nezávisle premenných. Celkovú premenlivosť môžeme teda hodnotiť pomocou SKO, ktorú môžeme napísať v tvare

$$s_Y^2 = s_y^2 + s_{y,x}^2 \quad (96)$$

V prípade, že nezávisle premenné nie sú korelované, prvý člen rovnice na pravej strane vyjadruje závislosť y na x. Čím je táto závislosť tesnejšia, väčšia, tým viac sa tento člen podieľa na celkovej hodnote. Druhý člen pravej strany rovnice vyjadruje vplyv ostatných nezávisle premenných na y.

V prípade, že nezávisle premenné sú navzájom korelované, neodrážajú členy rovnice (96) v čistej forme príslušné vplyvy. Ukazovatele tesnosti korelačnej závislosti ktoré uvedieme v ďalšom, v tomto prípade neurčujú veľkosť korelačného vzťahu, ale iba spoľahlivosť regresného odhadu – určenia regresnej funkcie. Potom prvý člen udáva, akú časť korelačného vzťahu regresnej funkcie ukazuje a druhý člen udáva, akú časť korelačného vzťahu môže regresná funkcia ukázať.

Korelačný index. Je to prvý ukazovateľ tesnosti závislosti y na x. Určíme ho zo vzťahu

$$I_{xy} = \sqrt{\frac{n \sum Y^2 - (\sum Y)^2}{n \sum y^2 - (\sum y)^2}} \quad (97)$$

kde n je rozsah súboru;  $\sum y^2$  je suma kvadrátov hodnôt premennej y;  $\sum Y^2$  je suma kvadrátov modelových hodnôt  $Y^2 = (a_{yx} + b_{yx} \cdot X)^2$ ;  $(\sum Y)^2$  je kvadrát sumy hodnôt závisle premennej y. Pri výpočte korelačného indexu z hodnôt triedených podľa x nám korelačný index udáva iba rozptyl prvého člena z rovnice (96).

Korelačný koeficient. Lineárna závislosť je najvýznamnejším prípadom korelačného vzťahu medzi dvomi veličinami. Korelačný index v tomto prípade nazývame korelačným koeficientom. Korelačný koeficient vypočítame podľa vzťahu

$$r_{yx} = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (98)$$

kde  $\sum xy$  je suma súčinu jednotlivých dvojíc hodnôt; ostatné členy ako v predošlom vzorci.

Medzi korelačným a regresným koeficientom platí vzťah

$$r_{yx} = b_{yx} \frac{s_x}{s_y} \quad (99)$$

Hodnota korelačného koeficientu sa pohybuje v intervale (-1;+1). Kladné znamienko značí **kladnú** a záporné **zápornú korelačnú závislosť**. Čím viacej sa bude hodnota  $r_{yx}$  blížiť k 1

alebo k  $-1$ , tým bude korelačný vzťah tesnejší. V prípade, že  $r_{yx}$  dosiahne hodnotu 1 alebo  $-1$ , pôjde o závislosť funkčnú. Čím viac sa bude  $r_{yx}$  blížiť k nule, tým bude korelačný vzťah slabší. V prípade, že  $r_{yx} = 0$ , budú obe premenné veličiny lineárne nekorelované (Obr.24).

Na veľkosť korelačného koeficienta vplyvajú mnohé faktory, ktoré nesúvisia s tesnosťou korelačného vzťahu. Z mnohých sú to hlavne:

**veľkosť jednotky** ako nositeľa korelovaných veličín; čím sú jednotky väčšie, tým sú tiež väčšie ukazovatele tesnosti korelačného vzťahu.

**extrémne hodnoty**; ich prítomnosť môže podstatne zväčšiť korelačný koeficient a tým podstatne skresliť predstavu o tesnosti korelačného vzťahu. Preto pri skúmaní korelačného vzťahu tieto hodnoty vylučujeme (dobrý príklad výpočtu korelačného koeficientu nájdeme napr. v [1] str.154).

V prípade, že nemáme možnosť určiť regresnú funkciu na určenie tesnosti korelačného vzťahu používame korelačný pomer. Korelačný pomer určujeme vždy z triednych údajov. Pre korelačný pomer platí vzťah

$$\eta_{yx} = \sqrt{1 - \frac{s^2}{s_y^2}} \quad (100)$$

kde  $s^2$  je SKO hodnôt  $x$  a  $y$ ;  $s_y^2$  je SKO  $y$  (podrobnejšie vid' [4] str.155).

Hodnota  $\eta_{yx}$  je závislá na triedení oboch premenných veličín. Čím je triedenie nezávislej premennej jemnejšie, tým je hodnota  $\eta_{yx}$  väčšia. Na druhej strane, pri danom triedení  $x$ , bude  $\eta_{yx}$  dávať tým lepšie výsledky, čím bude triedenie  $y$  jemnejšie.

Koeficient determinácie. V niektorých prípadoch sa používa aj koeficient determinácie, ktorý nadobúda hodnoty (0, 1) a je to vlastne mocnina korelačného koeficientu

$$r_{yx}^2 \quad (101)$$

Tento koeficient, ak násobíme číslom 100, udáva v percentách podiel zmien nezávislej premennej  $x$  na rozptyle závislej premennej  $y$ .

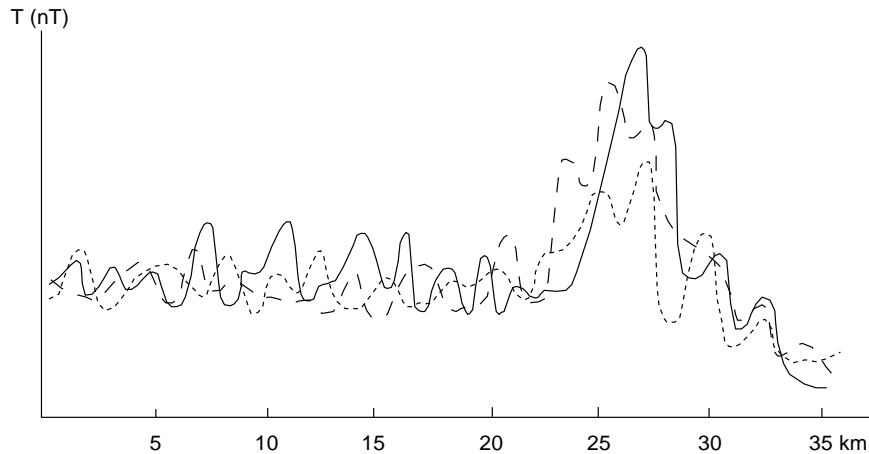
## 9. Náhodné funkcie

Teória náhodných funkcií má pre geológiu dôležitý význam. V praxi sa vlastne stále narába s náhodnými funkciami. V ďalšom si uvedieme pojem náhodnej funkcie a jej základné charakteristiky.

### 9.1 Pojem náhodnej funkcie

Doteraz sme sa zaoberali náhodnými veličinami. Náhodná veličina je charakteristická tým, že ako výsledok pokusu, nadobúda predtým neznámu ale určitú jednu hodnotu. Jednako, v určitých prípadoch nám takéto schéma pri štúdiu náhodného javu nestačí. Náhodná veličina sa v procese pokusu mení. Napríklad chod gravimetra. Takúto náhodnú veličinu nazývame **náhodná funkcia**. Náhodná funkcia sa nazýva takéto funkcia, ktorá ako výsledok pokusu, môže nadobudnúť dopredu neznámy tvar.

Konkrétny tvar, ktorý náhodná funkcia ako výsledok pokusu nadobudne, sa nazýva **realizácia náhodnej funkcie**. Ak nad náhodnou funkciou prevedieme niekoľko pokusov, získame súbor náhodných funkcií (obr.14). Ako každá funkcia, tak aj náhodná funkcia sa vyjadruje v tvare  $X(t)$ . Hodnota náhodnej funkcie pre určitú hodnotu argumentu  $t$  sa nazýva **prierez náhodnej funkcie**.



**Obr.27.** Výsledky opakovaných magnetometrických meraní

Je zrejmé, že náhodná funkcia môže mať viac argumentov – ak bude funkciou polohy, tak v rovine dva, v priestore tri a pod. V ďalšom sa budeme zaoberať náhodnou funkciou jedného argumentu.

Tak ako náhodná veličina, aj náhodná funkcia má svoje rozdelenie. K názornej predstave rozdelenia náhodnej funkcie sa dostaneme touto úvahou: hodnota prierezu náhodnej funkcie v bode  $t$  je  $X(t)$ . Je zrejmé, že táto hodnota je náhodnou veličinou, ktorá má svoje rozdelenie, závislé na  $t$ . Označme ho  $f(x,t)$ . Funkciu  $f(x,t)$  nazývame **jednorozmerným zákonom rozdelenia náhodnej funkcie**  $X(t)$ . Funkcia  $f(x,t)$  nie je vyčerpávajúcou charakteristikou náhodnej funkcie  $X(t)$  – charakterizuje iba zákon rozdelenia  $X(t)$  pre dané  $t$ . Neodráža závislosť  $X(t)$  na rôznych hodnotách  $t$ . Potom lepšou charakteristikou je dvojrozmerný zákon rozdelenia

$$f(x_1, x_2; t_1, t_2) \tag{102}$$

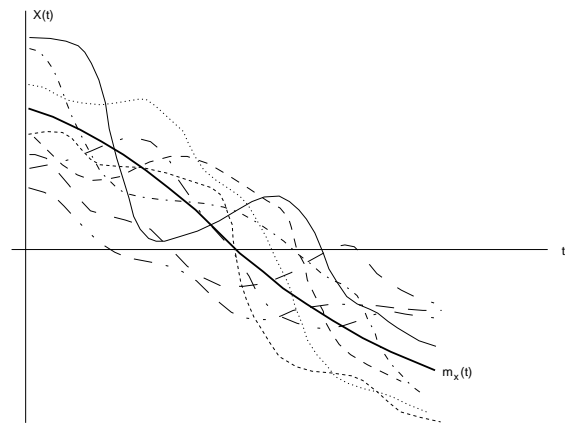
Analogicky, vyčerpávajúcou charakteristikou náhodnej funkcie by bol  $n$ -rozmerný zákon rozdelenia pre  $n \rightarrow \infty$ . Tento postup je však pre praktické použitie nepoužiteľný, preto sa v ďalšom budeme zaoberať jednoduchými formami charakteristiky náhodnej funkcie, analogickými náhodnej veličine.

## 9.2 Charakteristiky náhodnej funkcie

Predtým než uvedieme niektoré charakteristiky náhodnej funkcie je potrebné upozorniť na nasledovnú vec: na rozdiel od číselných charakteristík náhodnej veličiny, ktoré predstavovali určité čísla, charakteristiky náhodnej funkcie predstavujú vo všeobecnom prípade nie čísla, ale **funkcie**.

Aritmetický priemer náhodnej funkcie určíme nasledovným spôsobom: určíme prierez  $X(t)$  v určitom bode  $t_k$ . V tomto priereze dostaneme obyčajnú náhodnú veličinu. Určíme jej aritmetický priemer. Je zrejmé, že vo všeobecnom prípade bude tento priemer závisieť od  $t$ , t.j. je funkciou  $t$ :

$$\bar{X}_x(t) = m[X(t)] \quad (103)$$



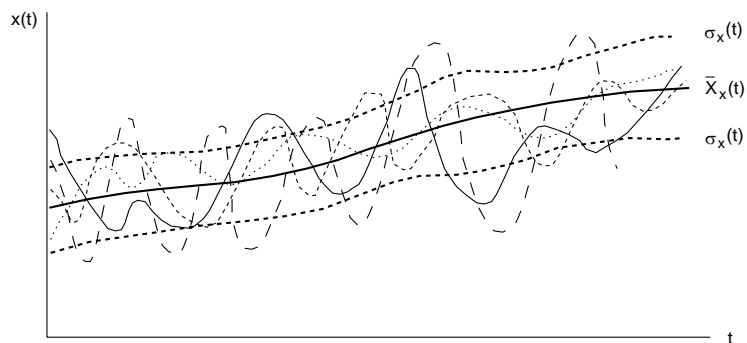
**Obr.28.** Aritmetický priemer NF

aritmetickým priemerom náhodnej funkcie  $X(t)$  nazývame náhodnú funkciu  $\bar{X}(t)$ , ktorá pri každej hodnote argumentu  $t$  je rovná hodnote aritmetického priemeru odpovedajúcejmu prierezu náhodnej funkcie.

To znamená, že aritmetický priemer náhodnej funkcie je určitá stredná funkcia, okolo ktorej rozličným spôsobom kolíšu konkrétne realizácie náhodnej funkcie.

Stredná kvadratická odchýlka náhodnej funkcie. Analogicky ako pre aritmetický priemer, možno aj pre SKONF náhodnej funkcie napísať: strednou kvadratickou odchýlkou náhodnej funkcie  $X(t)$  nazývame náhodnú funkciu  $\sigma_x(t)$ , ktorej hodnota pre každé  $t$  je rovná SKO odpovedajúcej prierezu náhodnej funkcie

$$\sigma_x(t) = \sigma[X(t)] \quad (104)$$



**Obr.29.** Stredná kvadratická odchýlka NF

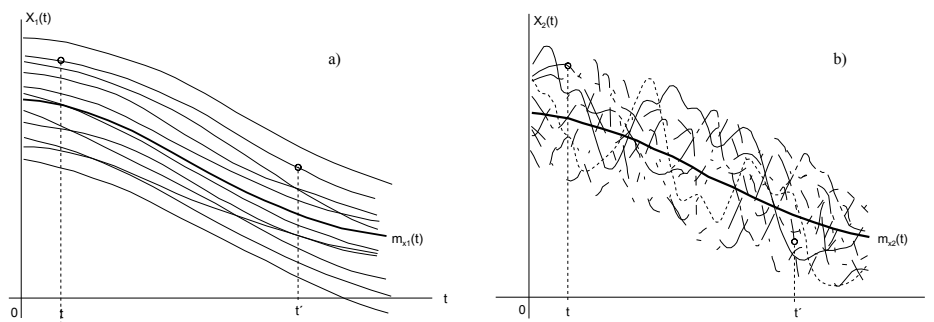
SKONF pri každom  $t$  charakterizuje rozptyl možných realizácií náhodnej funkcie voči strednej, t.j. charakterizuje stupeň náhodnosti náhodnej funkcie.

Analogicky pre disperziu náhodnej funkcie platí

$$\sigma_x^2(t) = \sigma^2[X(t)] \quad (105)$$

Korelačná funkcia. Pozrime sa na obr.30a a 30b, ktoré zobrazujú dve náhodné funkcie. Obe náhodné funkcie majú rovnaké aritmetické priemery i SKO. Na prvý pohľad sa však obe náhodné funkcie líšia. Vidíme, že aritmetický priemer a SKO nestačia na charakterizovanie náhodnej funkcie.





**Obr.30.** Grafy realizácií náhodných funkcií  $X_1(t)$  a  $X_2(t)$ .

Aby bolo možné uvedený jav charakterizovať, zavádza sa tzv. **korelačná funkcia**. Z obr.30 je zrejmé, že realizácia funkcie  $X_1(t)$  má v bodoch  $t$  a  $t'$  rovnakú vzdialenosť od funkcie  $\bar{X}_1(t)$ . Naproti tomu každá z realizácií funkcie  $X_2(t)$  má v bodoch  $t$  a  $t'$  náhodnú vzdialenosť od funkcie  $\bar{X}_2(t)$ . Vzťah medzi veličinami  $X(t)$  a  $X(t')$  vyjadríme pomocou **korelačného momentu**  $K(t,t')$ , ktorý závisí na polohe bodov  $t$  a  $t'$ . Každému páru hodnôt  $t$  a  $t'$  patrí určitá hodnota korelačného momentu. Potom **korelačnou funkciou** nazývame náhodnú funkciu dvoch argumentov  $t$  a  $t'$  -  $K(t,t')$ , ktorá pre každý pár hodnôt  $t$  a  $t'$  je rovná korelačnému momentu im odpovedajúcich hodnôt  $X(t)$  a  $X(t')$

$$K_x(t,t') = \overline{X(t)X(t')} \quad (106)$$

kde

$$\dot{X}(t) = X(t) - \bar{X}_x(t)$$

a 
$$\dot{X}(t') = X(t') - \bar{X}_x(t')$$

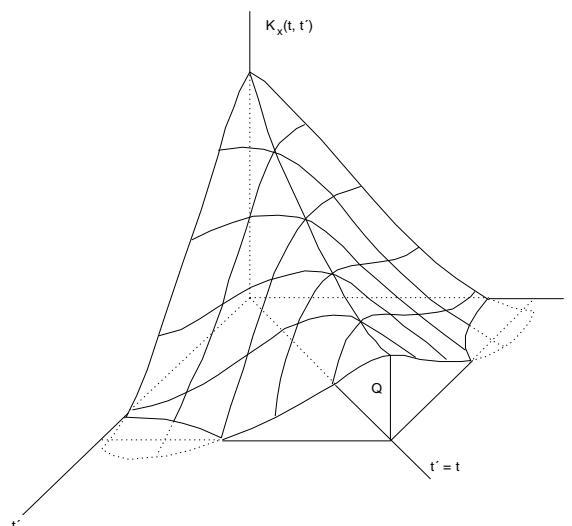
Priebeh korelačného momentu je možné graficky zobrazit' pomocou **normovanej korelačnej funkcie**  $r_x(t,t')$  v systéme súradníc  $t$  a  $t'$

$$r_x(t,t') = \frac{K_x(t,t')}{\sigma_x(t)\sigma_x(t')} \quad (107)$$

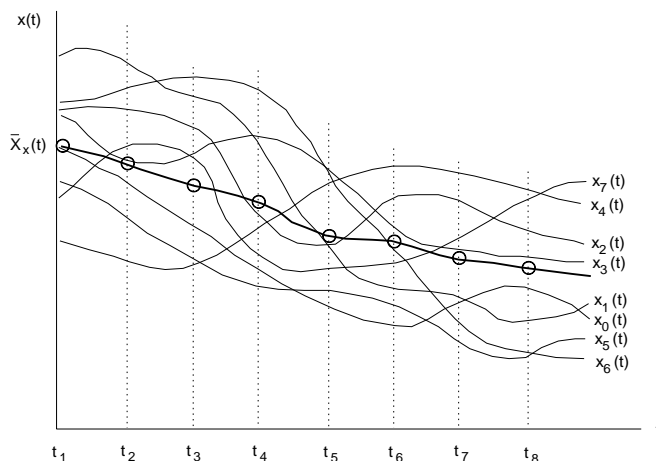
Príklad zostrojenia grafu normovanej korelačnej funkcie je v práci [2]. Pozn.: pre  $t = t'$  je

$$K_x(t,t') = r_x(t,t') = \sigma_x^2(t) = 1 \quad (108)$$

**Obr.32.** K výpočtu charakteristík NF



**Obr.31.** Graf normovanej korelačnej funkcie



### 9.3 Výpočet charakteristík náhodnej funkcie

Predpokladajme, že pre určitú náhodnú funkciu máme pokusom zistených niekoľko realizácií. Treba nájsť tieto charakteristiky náhodnej funkcie:

- aritmetický priemer -  $\bar{X}_x(t)$
- strednú kvadratickú odchýlku -  $\sigma_x^2(t)$
- korelačnú funkciu -  $K_x(t, t')$

Aby sme mohli túto úlohu splniť, realizujeme rad prierezov náhodnej funkcie (obr.32). V riadkoch sú hodnoty jednotlivých realizácií a v stĺpcoch sú ich prierezy.

**Tab.4.** Tabuľka pre výpočet štatistických charakteristík NF

	$t_1$	$t_2$	...	$t_k$	...	$t_l$	...	$t_m$
$X_1(t)$	$X_1(t_1)$	$X_1(t_2)$	...	$X_1(t_k)$	...	$X_1(t_l)$	...	$X_1(t_m)$
$X_2(t)$	$X_2(t_1)$	$X_2(t_2)$	...	$X_2(t_k)$	...	$X_2(t_l)$	...	$X_2(t_m)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X_i(t)$	$X_i(t_1)$	$X_i(t_2)$	...	$X_i(t_k)$	...	$X_i(t_l)$	...	$X_i(t_m)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X_n(t)$	$X_n(t_1)$	$X_n(t_2)$	...	$X_n(t_k)$	...	$X_n(t_l)$	...	$X_n(t_m)$

Z hodnôt v tabuľke potom nasledovným spôsobom vypočítame jednotlivé charakteristiky:

Aritmetický priemer

$$\bar{X}_x(t_k) = \frac{1}{n} \sum_{i=1}^n x_i(t_k) \quad (109)$$

Strednú kvadratickú odchýlku

$$\sigma_x(t_k) = \left[ \left( \frac{1}{n} \sum_{i=1}^n [x_i(t_k)]^2 - [\bar{X}_x(t_k)]^2 \right) \frac{n}{n-1} \right]^{1/2} \quad (110)$$

Korelačnú funkciu

$$K_x(t_k, t_l) = \left[ \frac{1}{n} \sum_{i=1}^n x_i(t_k) x_i(t_l) - \bar{X}_x(t_k) \bar{X}_x(t_l) \right] \frac{n}{n-1} \quad (111)$$

## Literatúra:

- [1] Reisenauer,R.: Metody matematické statistiky a jejich aplikace v technice. SNTL, Praha 1970.
- [2] Ryžov,P.A.,Gutkov,V.M.: Primenenie matematickej statistiky pri razvedke neodr. Moskva 1966.
- [3] Marek,F. a kol.: Magnetometrie I. Praha 1968.
- [4] Egermayer,F., Kaňok,M.: Elementární statistické metody v průmyslové výrobě. SNTL/SVT:. Praha 1965
- [5] Ventcel,E.S.: Teorija verojatnostej. Nauka, Moskva 1969.
- [6] Doerfel,K.: Statistika v analitičeskoj chimii. Mir, Moskva 1969
- [7] Janák,F., Kadlec,E.: Statistické spracování fyzikálních parametrů s aplikací na horninách Spišsko-gemerského rudohorí. Výsk. zpráva, Geofyzika Brno 1969.
- [8] Michajlova,N.P.: Po povodu ispolzovanija matematickoj statistiky pri izučenii magnitnych svojstv. Mat.konf., Kijev 1971.
- [9] Brjuzgina,H.I.: Primenenije disperznogo analiza pri petrofizičeskyh issledovanijach. Geologija i geofyzika. No.8, 1970.
- [10] Fisher,R.A.: Dispersion on a sphere. Proc.roy.soc., London, A, 217, p.295-305, 1953.