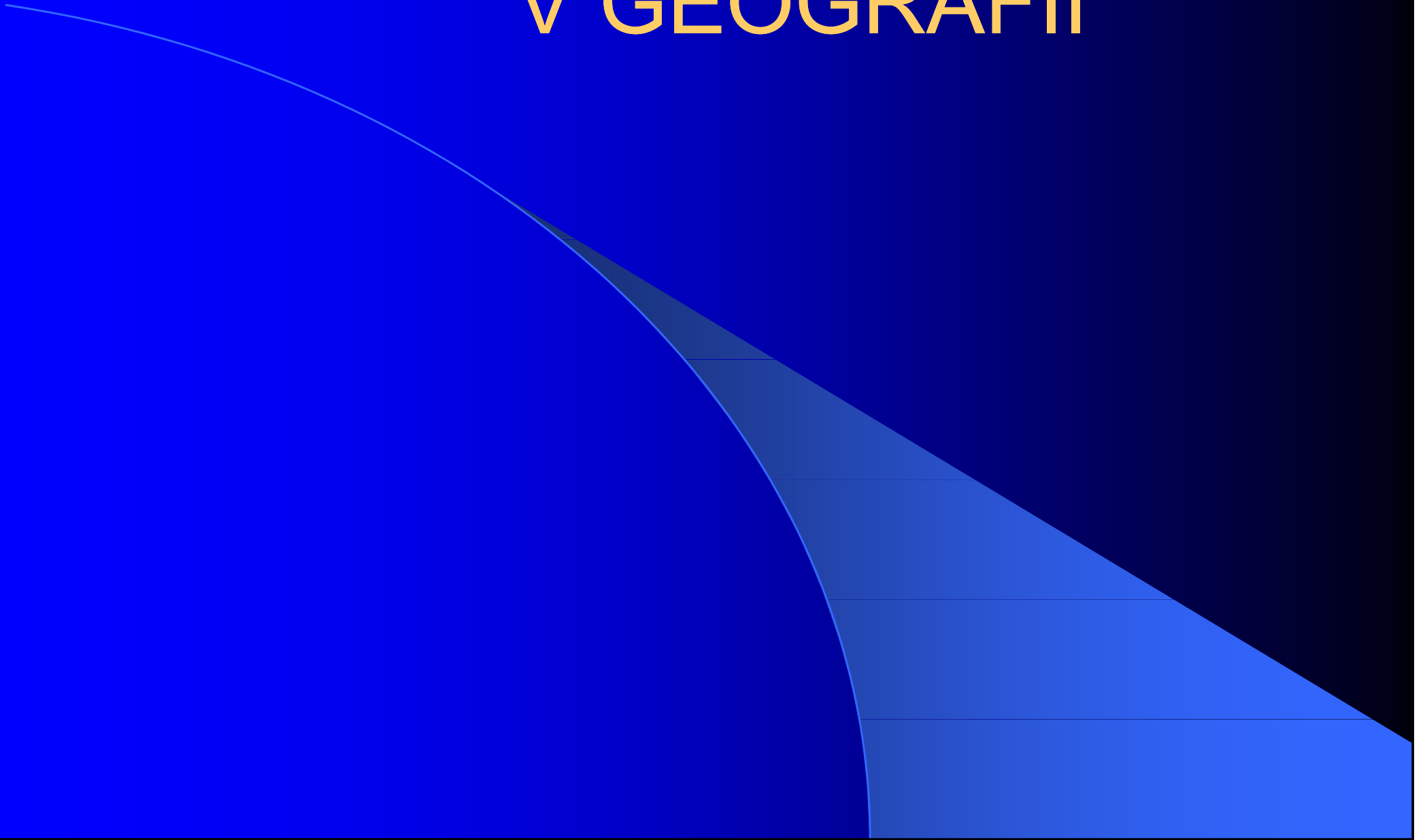


STATISTICKÉ METODY V GEOGRAFII



STATISTICKÉ METODY V GEOGRAFII



Karl Friedrich
Gauss
1777-1855

Teoretická rozdělení

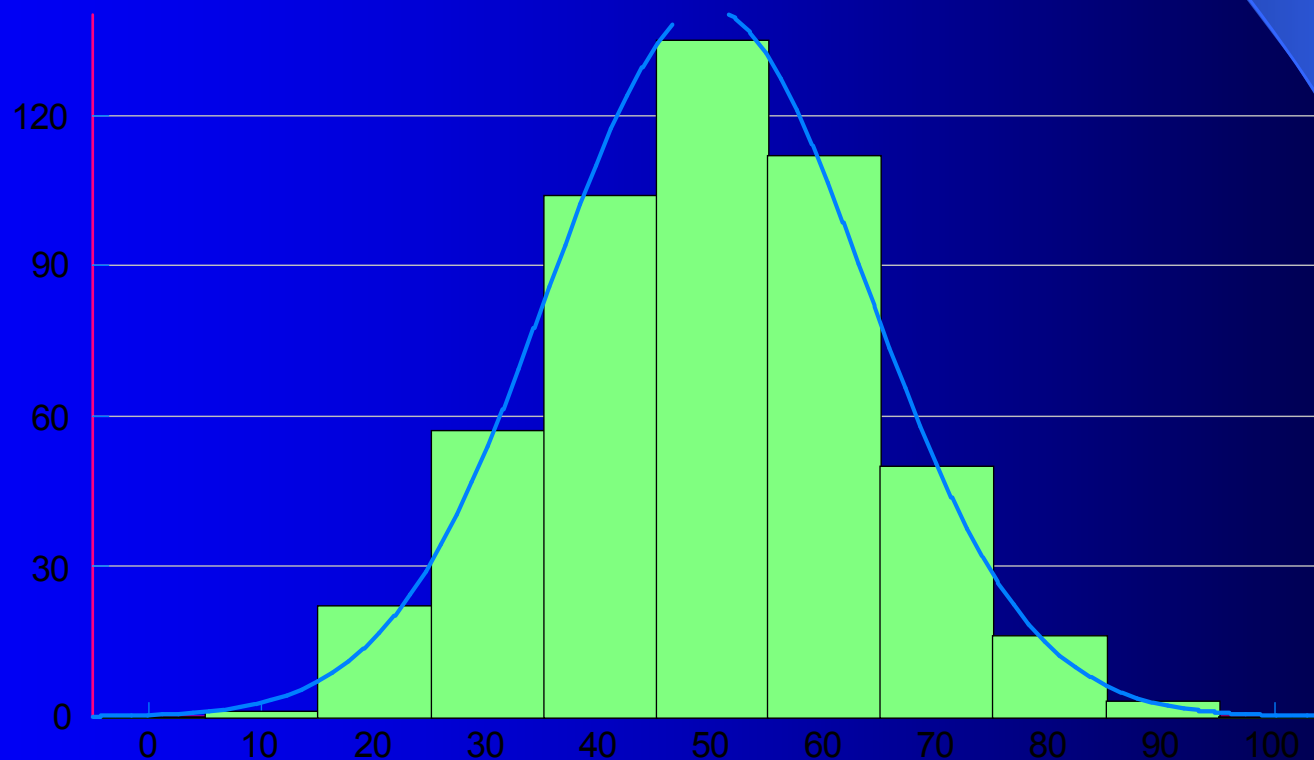
Základní pojmy

- náhodná veličina **spojitá**
- Může teoreticky nabývat nekonečného množství hodnot z určitého intervalu např. teplota)

- náhodná veličina **nespojité**
- Nabývá jen konečného množství hodnot urč. Intervalu. Např. počet měsíců s teplotou nad...)
- Každé hodnotě je možno přiřadit pravděpodobnost jejího výskytu, součet všech dílčích pravděpodobností je 1

Teoretická rozdělení

- histogram – grafické znázornění četností
- rozsah souboru se blíží k nekonečnu + náhodná veličina je spojitá
- – frekvenční funkce / hustota pravděpodobnosti

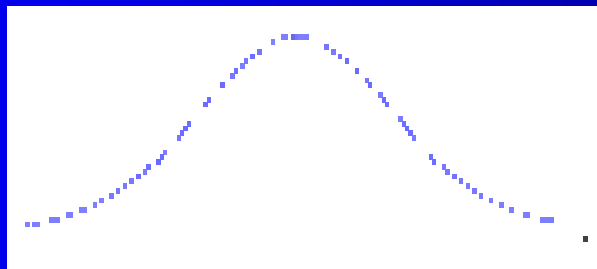


- kumulativní relativní četnost tj. součtová čára -
- distribuční funkce
- obr.

Normální rozdělení / Gaussovo, Laplaceovo- Gaussovo

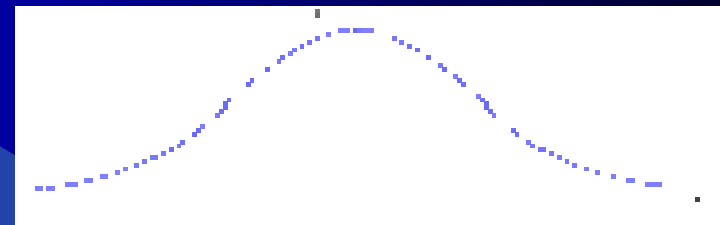
- Normální rozdělení se univerzálně používá k aproximaci (k přibližnému vyjádření) rozdělení pravděpodobnosti velkého množství náhodných veličin (v biologii, technice, ekonomii atd.)

Hustota pravděpodobnosti normálního rozdělení je symetrická zvonovitá **Gaussova křivka**.

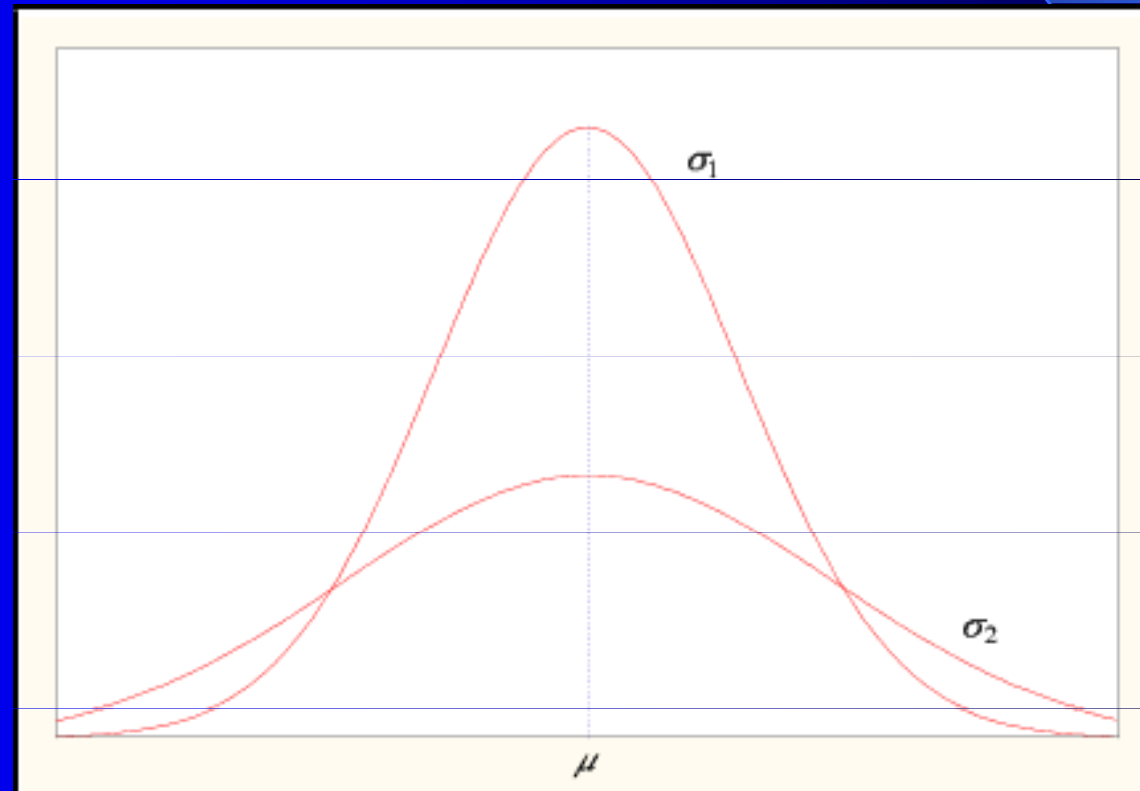


Normální rozdělení

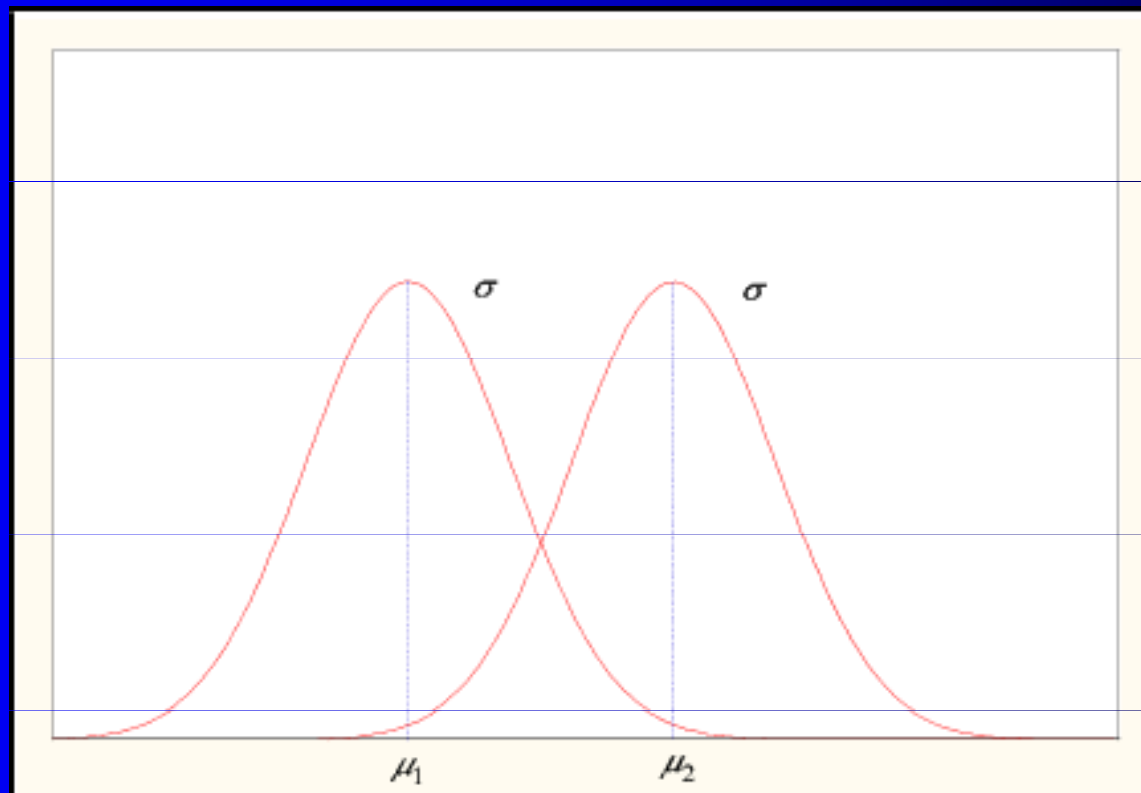
- Zvonovitý tvar
- Souměrný
- Šikmost 0, špičatost 0
- Asymptoticky se blíží 0



- Normální rozdělení s parametry:
- stejný průměr, různé směrodatné odchylky
- čím větší odchylka , tím „plošší tvar rozdělení

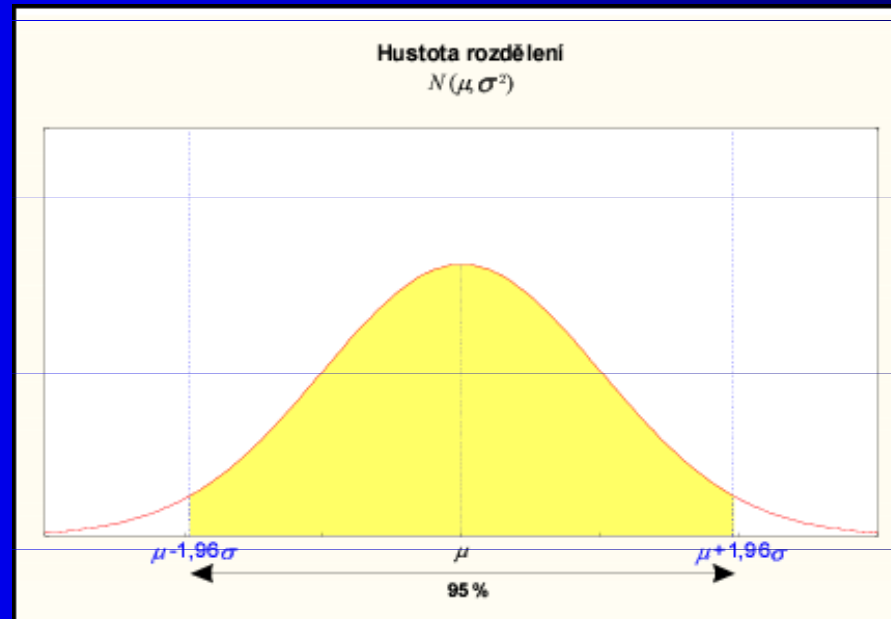


- Normální rozdělení
- různé průměry, stejná směrodatná odchylka



Normální rozdělení / Gaussovo pokračování

- Normální křivka a osa x vymezují plochu 100%,
- tj. lze stanovit pravděpodobnosti, s nimiž leží hodnoty v určitém intervalu,
- hranice intervalu tvoří průměr a násobky směrodatné odchylky
- obr.



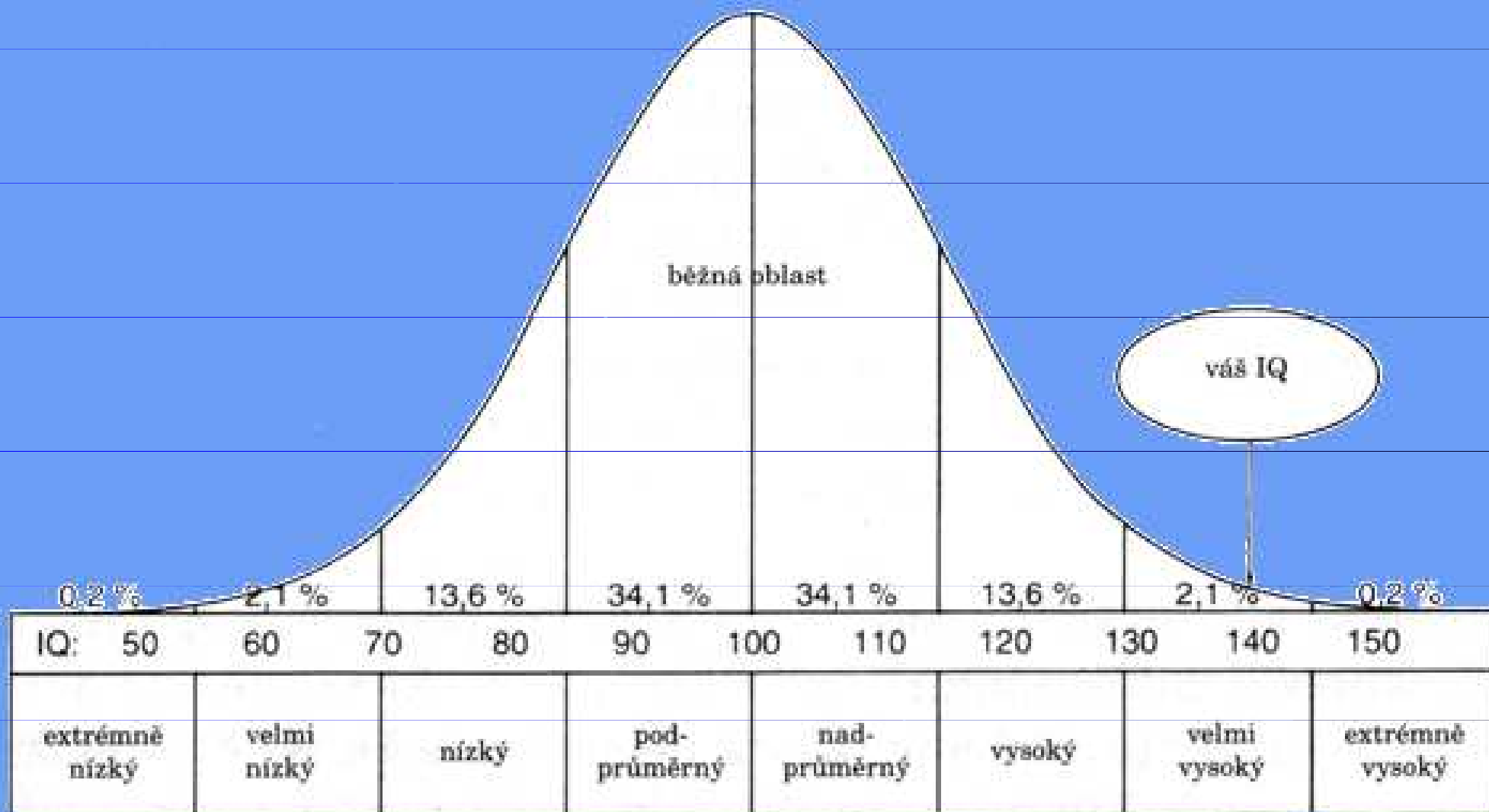
V normálním rozdělení:

- **68, 27% leží v intervalu:**
- **(průměr + - směr. odchylka)**

- **95% leží v intervalu:**
- **(ar. průměr +- 1,96 směr. odchylky)**

- **99% leží v intervalu:**
- **(ar. průměr +- 2,576 směr. odchylky)**

Normální rozdělení pro IQ



imbecilita

debilita

Lehká d.

průměr

vynikající

genialita

idocie

IQ (v bodech)	stupeň inteligence případů (v %)	procento zkoumaných
méně než 20	idiocie	0,1
20 - 49	imbecilita	0,5
50 - 69	debilita	1,9
70 - 79	tzv. lehká debilita	5,0
80 - 89	podprůměrná	14
90 - 109	průměrná	48
110 - 119	nadprůměrná	18
120 - 139	vynikající	11
140 a více	genialita	1,5



Příklady

Př.1

- Populace má v daném testu průměr 100, směrodatnou odchylku 15.
- Vypočítejte hranice intervalů, v kterém se nachází 68 % populace.

Příklad

- Výška v populaci chlapců ve věku 3,5 - 4 roky má normální rozdělení s průměrem 102 cm a směrodatnou odchylkou 4,5 cm.
- Vypočítejte hranice intervalu hodnot výšky , ve kterých se nachází
- A) 70%
- B) 95%
- C) 99%
- příslušné populace

Příklad 3

- zadání:
- Výška v populaci chlapců ve věku 3,5 - 4 roky má normální rozdělení s průměrem 102 cm a směrodatnou odchylkou 4,5 cm.
- Spočtete, jaké procento chlapců v uvedeném věku má výšku menší nebo rovnou 93 cm.

Řešení 3

- Pravděpodobnost, že výška nabude hodnoty menší nebo rovné 93 cm, je vyjádřena hodnotou **distribuční funkce F(93)** pro **parametry normálního rozdělení 102;4,5**

Microsoft Excel

Soubor Úpravy Zobrazit Vložit Formát Nástroje Data Okno Nápověda

NORMDIST = =NORMDIST(93;102;4.5;pravda)

NORMDIST

X	93	= 93
Střed_hodn	102	= 102
Sm_odch	4,5	= 4,5
Součet	pravda	= PRAVDA

= 0,022750062

Vrátí hodnotu normálního součtového rozdělení pro zadanou střední hodnotu a směrodatnou odchylku.

Součet je logická hodnota: součtová distribuční funkce = PRAVDA, hromadná pravděpodobnostní funkce = NEPRAVDA.

Výsledek = 0,022750062 = 2,27%

OK Storno

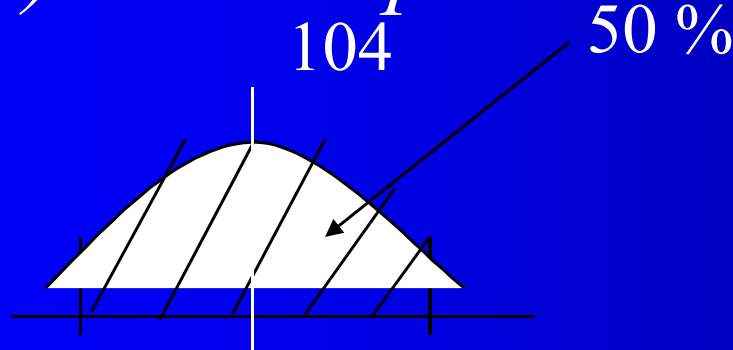
Odpověď: 2,27 % chlapců ve věku 3,5 – 4 roky je menších než 93 cm

Příklad 4

- Psychologickými testy bylo zjištěno, že hodnota IQ populace je náhodnou veličinou s normálním rozdělením, jehož střední hodnota je 104 a směrodatná odchylka 8.
- Určete hodnotu IQ, kterou podle uvedených pravděpodobnostních předpokladů:
 - meze, ve kterých bude 50% populace,
 -

Řešení 4

- a) *meze pro 50 % mužské populace*



Hledáme dolní a horní meze intervalu (hodnot IQ),
ve které se bude nacházet 50% mužské populace, tj 1. a 3. kvartil

Řešení 2a)

Excel, statistická funkce inverzní k e Gauss. - NORMINV

Microsoft Excel

Prst: 0,25 = 0,25
Střední: 104 = 104
Sm_odch: 8 = 8

= 98,60407707

Vrátí inverzní funkci k distribuční funkci normálního součtového rozdělení pro zadanou střední hodnotu a směrodatnou odchylku.
Prst je pravděpodobnost normálního rozdělení, číslo mezi 0 a 1 včetně.

Výsledek = 98,60407707

OK Storno

Microsoft Excel

Prst: 0,75 = 0,75
Střední: 104 = 104
Sm_odch: 8 = 8

= 109,3959229

Vrátí inverzní funkci k distribuční funkci normálního součtového rozdělení pro zadanou střední hodnotu a směrodatnou odchylku.
Prst je pravděpodobnost normálního rozdělení, číslo mezi 0 a 1 včetně.

Výsledek = 109,3959229

OK Storno

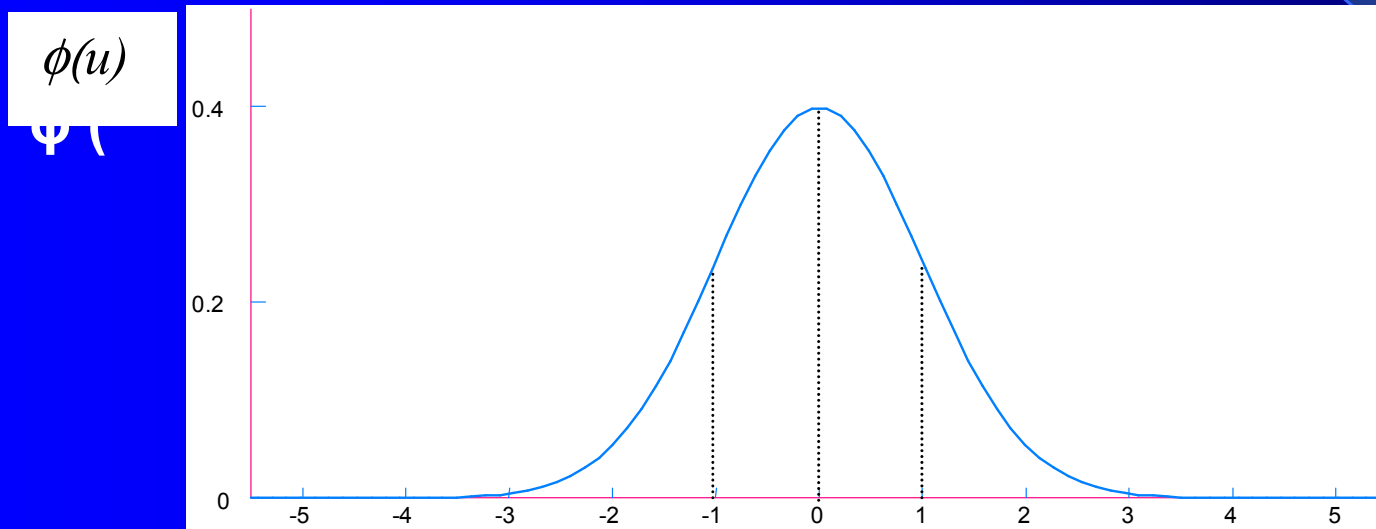
Podle parametrů daného normálního rozdělení 50 populace má IQ v intervalu 98,6 a 109,4.

- Pro normované normální rozdělení zavedeme označení $N(0, 1)$.

Normování hodnoty: od hodnoty se odečte aritmetický průměr,

výsledek (tj. odchylka) se dělí směr. odchylkou

Hustota pravděpodobnosti normovaného normálního rozdělení:



Tabulkové vyjádření vybraných hodnot hustoty pravděpodobnosti

u	0,0	0,5	1,0	1,5	2,0	2,5	3,0	3,5
$\phi(u)$	0,399	0,352	0,242	0,130	0,054	0,018	0,004	0,001

u

Tabulkové vyjádření vybraných hodnot distribuční funkce

u	0,0	0,5	1,0	1,5	2,0	2,5	3,0	3,5
$\Phi(u)$	0,500	0,691	0,841	0,933	0,977	0,994	0,999	0,999

Binomické rozdělení



Binomické rozdělení

- pro diskrétní náhodné proměnné,
- které mohou nabývat pouze dvou hodnot (např. ano, ne)
- pravděpodobnost, že nastane alternativa ANO označme π
- pravděpodobnost, že nastane NE ... $q = 1 - \pi$), protože
- platí $\pi + q = 1$ (100 %)
- k výpočtu se používá binomický rozvoj

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$P(X = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}, \quad \text{pro } k = 0, 1, 2, \dots, n.$$

Příklad 1 – binomické rozdělení

- Předpokládejme, že pravděpodobnost narození dívky je 0,49.
- Jaká je pravděpodobnost toho, že mezi třemi dětmi v rodině je právě jedna dívka?

Řešení 1

Tabulka3: Parametry binomického rozdělení v příkladu

Pokus	Úspěch	Neúspěch	Pravděpodobnost úspěchu	Počet pokusů	Počet úspěchů
				n	k
narození dítěte	dívka	chlapec	0,49	počet dětí	počet dívek

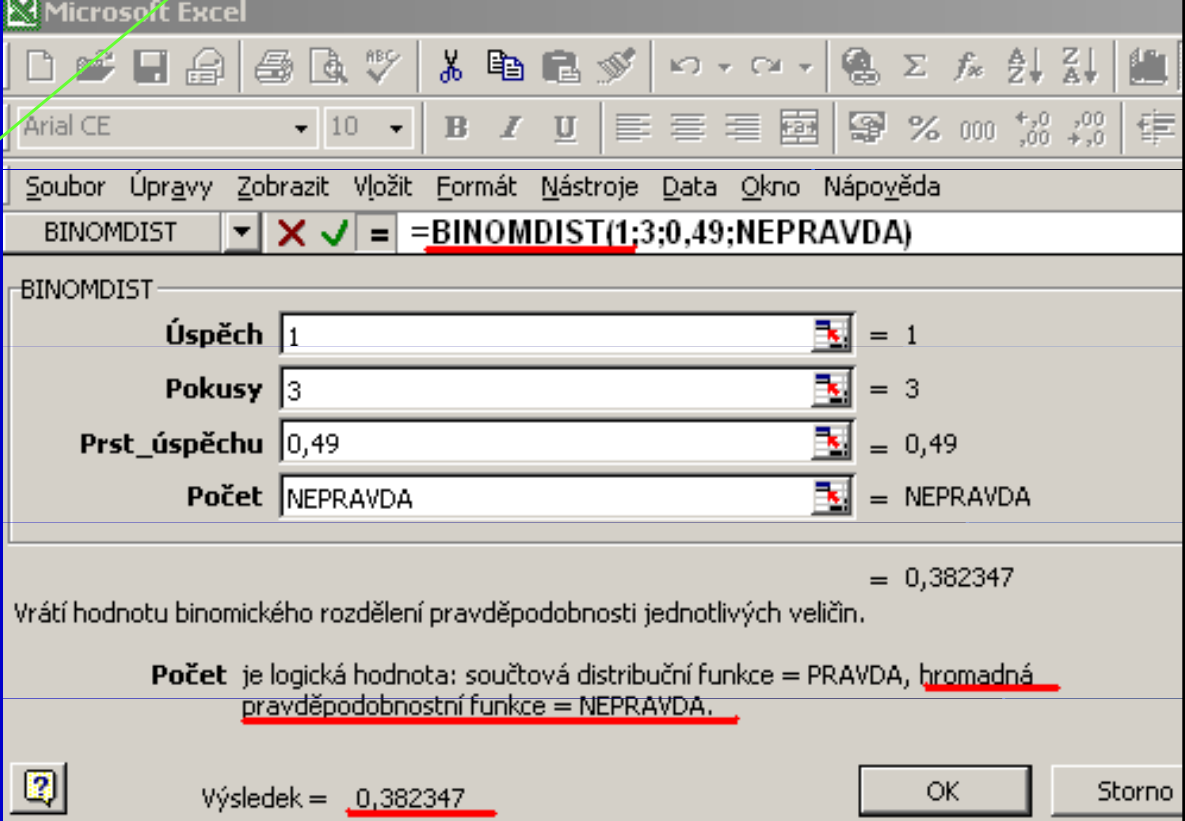
Řešení 1

Jak je vidět z tabulky, počet narozených dívek v rodině je náhodná veličina s binomickým rozdělením. Pravděpodobnost, že mezi třemi dětmi je právě jedna dívka, tedy vypočteme jako

$$P(X = 1) = \binom{3}{1} 0,49^1 \cdot 0,51^2 = 3 \cdot 0,127 = 0,38. \diamond$$

$$\binom{3}{1} = \frac{3!}{1!2!} = \frac{3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 1} = 3.$$

Pravděpodobnost, že ze tří dětí bude jedna dívka, je 38%.



Microsoft Excel

Binomická distribuce

Úspěch 1 = 1

Pokusy 3 = 3

Prst_úspěchu 0,49 = 0,49

Počet NEPRAVDA = NEPRAVDA

Výsledek = 0,382347

Vrátí hodnotu binomického rozdělení pravděpodobnosti jednotlivých veličin.

Počet je logická hodnota: součtová distribuční funkce = PRAVDA, hromadná pravděpodobnostní funkce = NEPRAVDA.

OK Storno

Příklad 2

Jaká je pravděpodobnost, že v rodině s 8 dětmi jsou právě 3 dívky? Pravděpodobnost narození dívky je 0,49.

Řešení

binomický rozvoj:

$$P(k = 3) = \binom{8}{3} 0,49^3 \cdot 0,51^5 = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} \cdot 0,118 \cdot 0,035 = 0,23. \diamond$$

Pravděpodobnost, že v rodině s 8 dětmi jsou tři dívky, je 0,23, tj. 23 %.

Příklad 2, binomické rozdělení

- Vypočítejte pravděpodobnost, se kterou se vyskytne určitý počet měsíců v roce hodnocených jako „suché“.
- Konkretizace:
 - oblast Oxford,
 - období 1851 – 1943, tj. 1116 měsíců
 - **Suchý měsíc** - tj. méně srážek v měsíci než je dlouhodobý průměr tohoto měsíce.
 - **617 měsíců hodnocených jako suché**
 - **499 – vlhké měsíce**

Řešení 2

„úspěch“	„neúspěch“	Pravděpodobnost suchého měsíce	Pravděpodobnost vlhkého měsíce	Počet měsíců	Počet suchých měsíců
suchý	vlhký	$\pi = 617/1116$ $\pi = 0,553$	$q = 499/1116$ $q = 0,447$ $(q = 1 - \pi)$	$n = 12$	$k = 0$ až 12

Řešení

- Ručně pomocí binomického rozvoje
- s podporou např. Excel

Řešíme dílčí příklady, tj. jaká je pravděpodobnost, že v roce se vyskytne

- žádný suchý měsíc, tj. $k = 0$
- Jeden suchý měsíc, tj. $k = 1$
- Atd.
- všechny měsíce suché, $k = 12$

Řešení 2

k	f(x)
0	0,000
1	0,000945
2	0,006428
3	0,026507
4	0,073785
5	0,146051
6	0,21
7	0,223
8	0,172
9	0,095
10	0,035
11	0,0079
12	0,0008

Microsoft Excel

Toolbar: File, Edit, Format, Tools, Data, Window, Help, etc.

Menu: Soubor Úpravy Zobrazit Vložit Formát Nástroje Data Okno Nápověda

Formula Bar: BINOMDIST = =BINOMDIST(5;12;0,553;nepravda)

BINOMDIST

Úspěch: 5 = 5

Pokusy: 12 = 12

Prst_úspěchu: 0,553 = 0,553

Počet: nepravda = NEPRAVDA

= 0,146050652

Vrátí hodnotu binomického rozdělení pravděpodobnosti jednotlivých veličin.

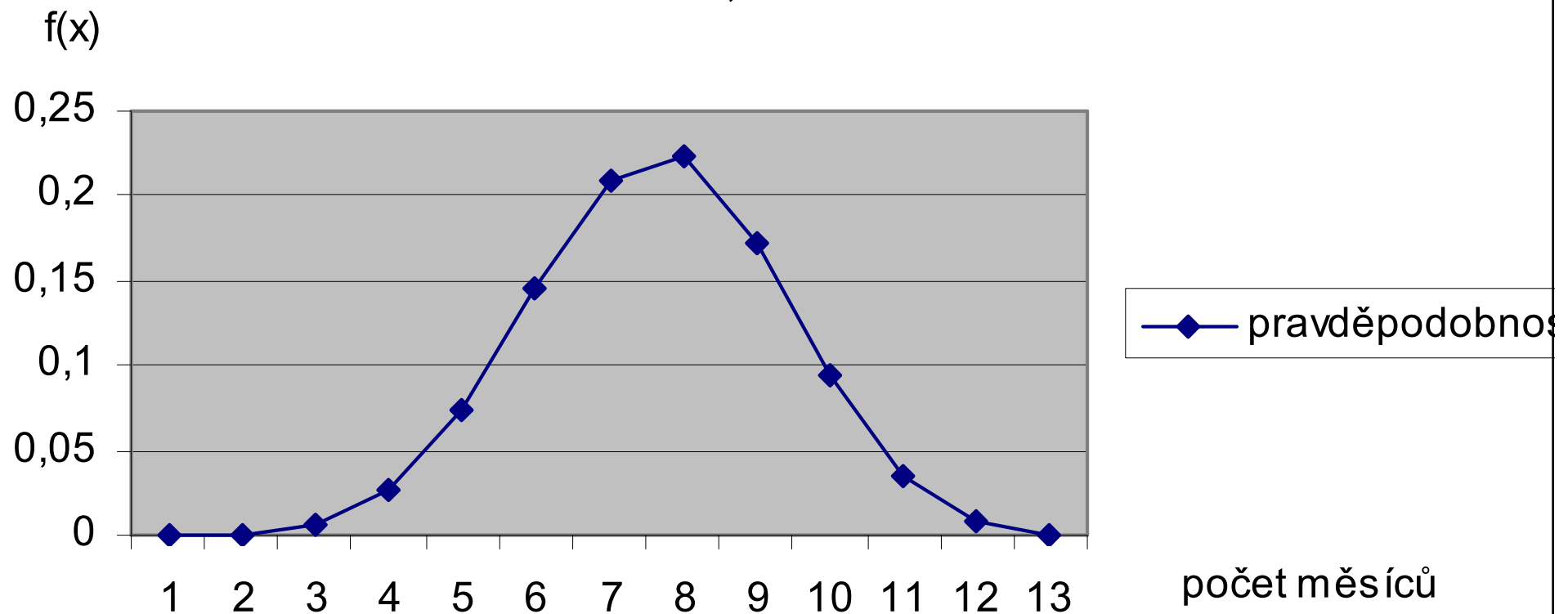
Úspěch je počet úspěšných pokusů.

Wrench icon

Výsledek = 0,146050652

OK Storno

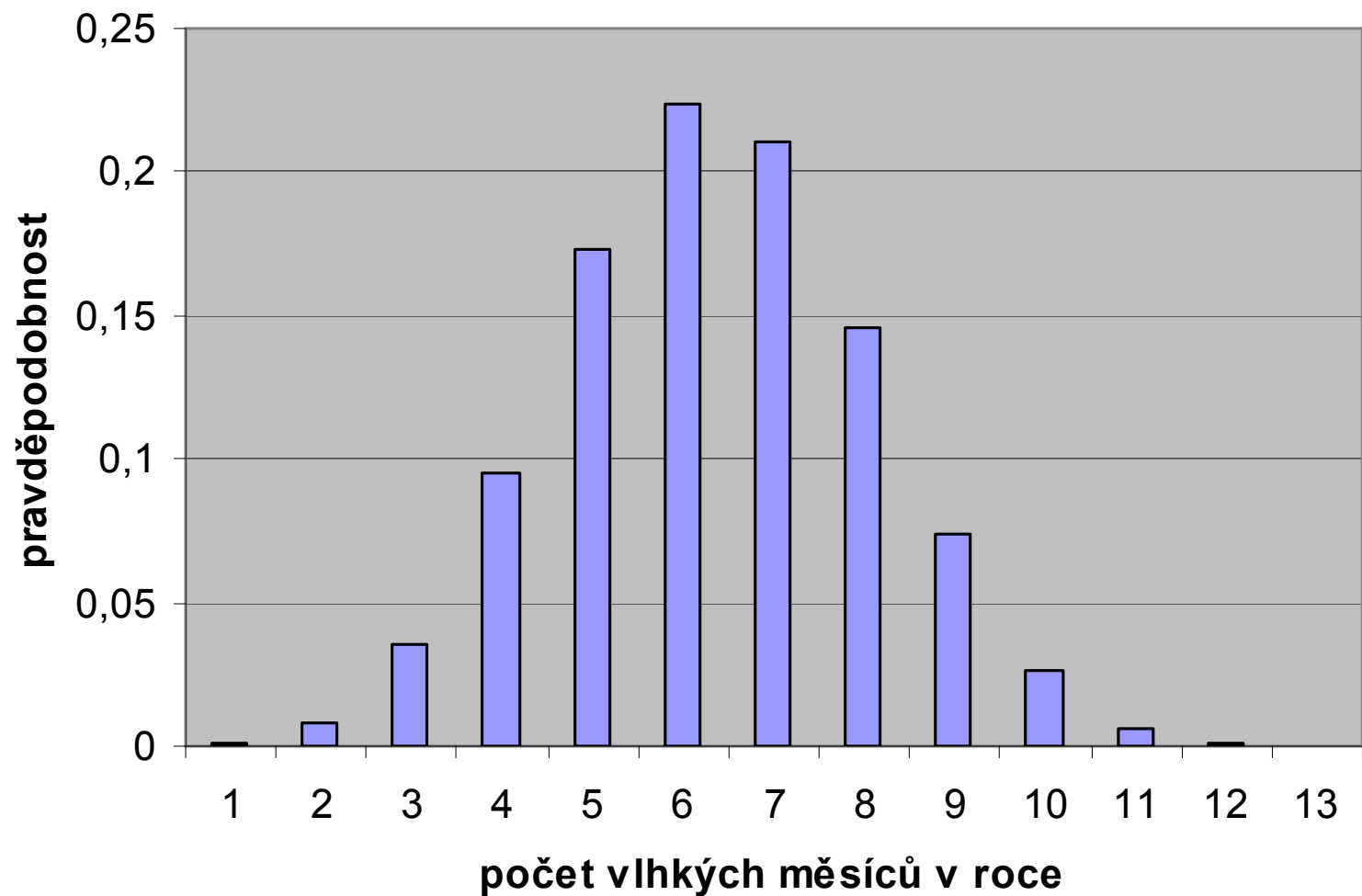
Pravděpodobnost počtu suchých měsíců v roce,
Oxford, 1851 - 1943



Jak bude vypadat situace pro „vlhke“ měsíce?

Binomické rozdělení

Pravděpodobnost výskytu vlhkého měsíce
v oblasti Oxfordu v letech 1851 - 1943



Poisson - příklad



Poissonovo rozdělení

- – pro rozdělení vzácných případů
- (zimní bouřka, výskyt mutace apod.).

- Je-li pravděpodobnost nějaké výjimečné události (např. určité mutace genu) relativně malá a rozsah výběru poměrně velký, pak **Poissonovo rozdělení v podstatě splývá s binomickým**, ale je mnohem výhodnější pro počítání .

Poisson - příklad

- Předpokládejme, že v určité populaci krys se vyskytuje albín s pravděpodobností
- $p = 0,001$, ostatní krys jsou normálně pigmentované.
- Ve vzorku 100 krys náhodně vybraných z této populace určete pravděpodobnost, že vzorek
 - a) neobsahuje albína,
 - b) obsahuje právě jednoho albína.

Řešení

- určete pravděpodobnost, že vzorek
- neobsahuje albína,

Microsoft Excel

Soubor Úpravy Zobrazit Vložit Formát Nástroje Data Okno Nápověda

BINOMDIST = =BINOMDIST(0;100;0,001;NEPRAVDA)

BINOMDIST

Úspěch	<input type="text" value="0"/>	= 0
Pokusy	<input type="text" value="100"/>	= 100
Prst_úspěchu	<input type="text" value="0,001"/>	= 0,001
Počet	<input type="text" value="NEPRAVDA"/>	= NEPRAVDA

= 0,904792147

Vrátí hodnotu binomického rozdělení pravděpodobnosti jednotlivých veličin.

Úspěch je počet úspěšných pokusů.

WYSIWYG Výsledek = 0,904792147

Pravděpodobnost, že neobsahuje albína, je 90,47 %

Řešení 3

Microsoft Excel

Binomická funkce: **BINOMDIST**

Úspěch 1 = 1

Pokusy 100 = 100

Prst_úspěchu 0,001 = 0,001

Počet NEPRAVDA = NEPRAVDA

= 0,090569784

Vrátí hodnotu binomického rozdělení pravděpodobnosti jednotlivých veličin.

Počet je logická hodnota: součtová distribuční funkce = PRAVDA, hromadná pravděpodobnostní funkce = NEPRAVDA.

Výsledek = 0,090569784

OK Storno

Pravděpodobnost, že 100 členná populace krys bude obsahovat albína, je 9 %.



Další rozdělení

Pearsonova křivka III. typu

- Na empirické rozdělení mnoha statistických souborů s nimiž v geografii pracujeme, **nelze aplikovat normální rozdělení.**
- Platí to například v těch případech, kdy studovaná náhodná veličina **nemá teoreticky zdůvodněnou možnost nabývat nekonečných hodnot** nebo je-li omezena konečnými čísly
V takovýchto případech lze aplikovat na studovaný soubor některou ze dvanácti křivek Pearsonova systému.

Pearsonova křivka III. typu

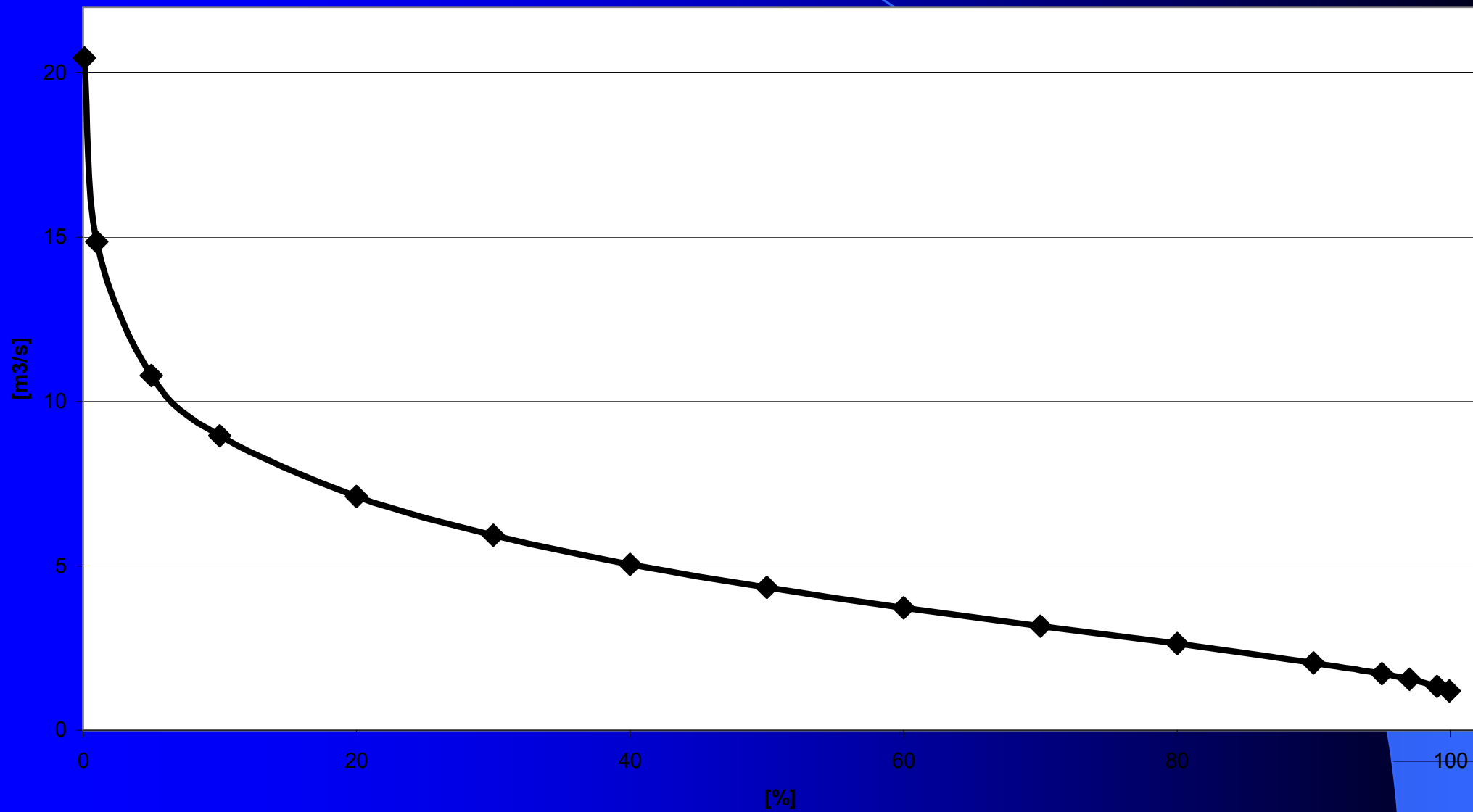
- Pearsonova křivka III. typu
- - obvykle pro veličiny s omezeným množstvím hodnot, které může nabývat
- - z křivky lze např. vyčíst pravděpodobnost se kterou bude hodnota sledovaného statistického znaku dosažena
- v hydrologii se počítá Pearsonova křivka ve variantě součtová čára četností jako
- tzv. čára překročení

- **příklad**

- Konstrukce čáry překročení z průměrných ročních průtoků vodního toku Lažánka za říjen 2002.

den	průtok Qd (m ³ /s)	den	průtok Qd (m ³ /s)
1	2,99	16	2,98
2	2,84	17	4,64
3	2,75	18	12,2
4	3,22	19	7,73
5	3,55	20	4,38
6	12,2	21	3,41
7	9,12	22	3,85
8	3,82	23	3,47
9	3,55	24	3,36
10	3,23	25	3,51
11	2,89	26	12,2
12	3,25	27	10,3
13	3,79	28	6,2
14	3,05	29	4,15
15	3,05	30	5,75
		31	5,1

Křivka překročení průměrných ročních průtoků vodního toku Lažánka za říjen 2002



rozdělení χ^2

- rozdělení χ^2 – náhodný výběr n prvků ze základního souboru (počet vybíraných prvků = počet stupňů volnosti)
- dostaneme n hodnot, součtu druhých mocnin daného počtu vybraných prvků odpovídá určitá křivka,

Studentovo/t/ rozdělení

- Studentovo/t/ rozdělení – hodnocení odchylek aritmetického průměru základního souboru a výběrových souborů, odchylkám přísluší Studentovo rozdělení