

DESKRIPCE

Na procvičení pro zkouškový test viz studijní materiály:

Cvičení 2018-05-02 (ukázky výpočtů do zkouškových testů)

(konkr. příklady 1-3)

Návod pro seminární práci následuje:

Charakteristiky číselných veličin

MEDIÁN (50% KVANTIL)

	A	B	C	D
1	pacientka č.	léčba	věk	počet dětí
2	1	3	27	0
3	2	1	42	3
4	3	2	36	2
				
20	19	1	46	2
21	20	3	28	1
22			=median(C2:C21)	

Charakteristiky číselných veličin

PRŮMĚR (AVERAGE; MEAN)

	A	B	C	D
1	pacientka č.	léčba	věk	počet dětí
2	1	3	27	0
3	2	1	42	3
4	3	2	36	2
				
20	19	1	46	2
21	20	3	28	1
22			=průměr(C2:C21)	

Charakteristiky číselných veličin

SMĚRODATNÁ ODCHYLKA

	A	B	C	D
1	pacientka č.	léčba	věk	počet dětí
2	1	3	27	0
3	2	1	42	3
4	3	2	36	2
				
20	19	1	46	2
21	20	3	28	1
22			35,9	1,55
			=smodch(C2:C21)	

Charakteristiky číselných veličin

KVARTILY pomocí EXCELU:

=QUARTIL(data;0) ... min

=QUARTIL(data;1) ... 25% kvantil

=QUARTIL(data;2) ... 50% kvantil

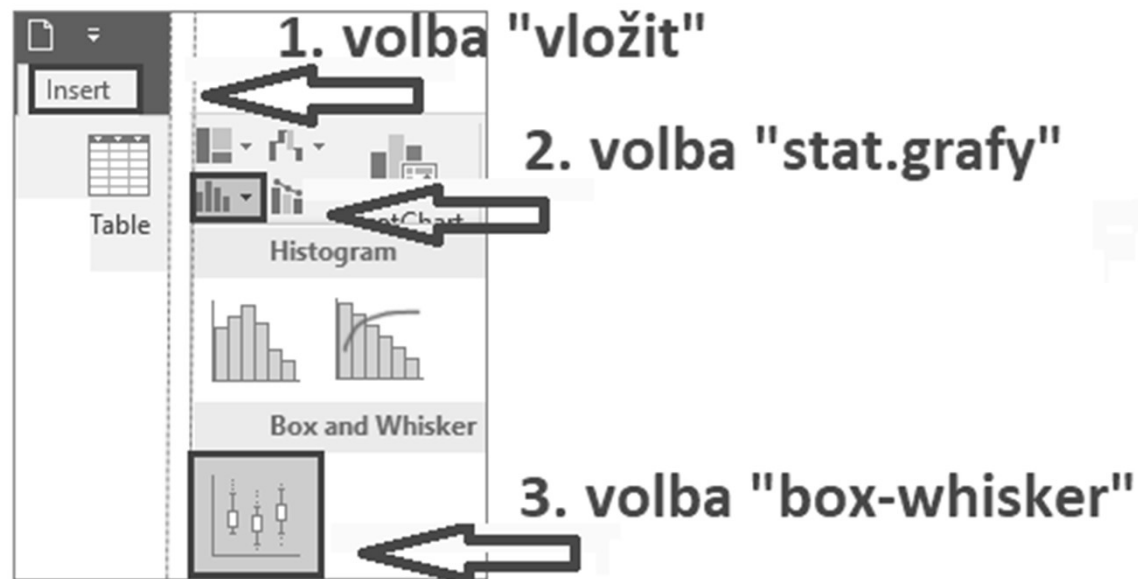
=QUARTIL(data;3) ... 75% kvantil

=QUARTIL(data;4) ... max

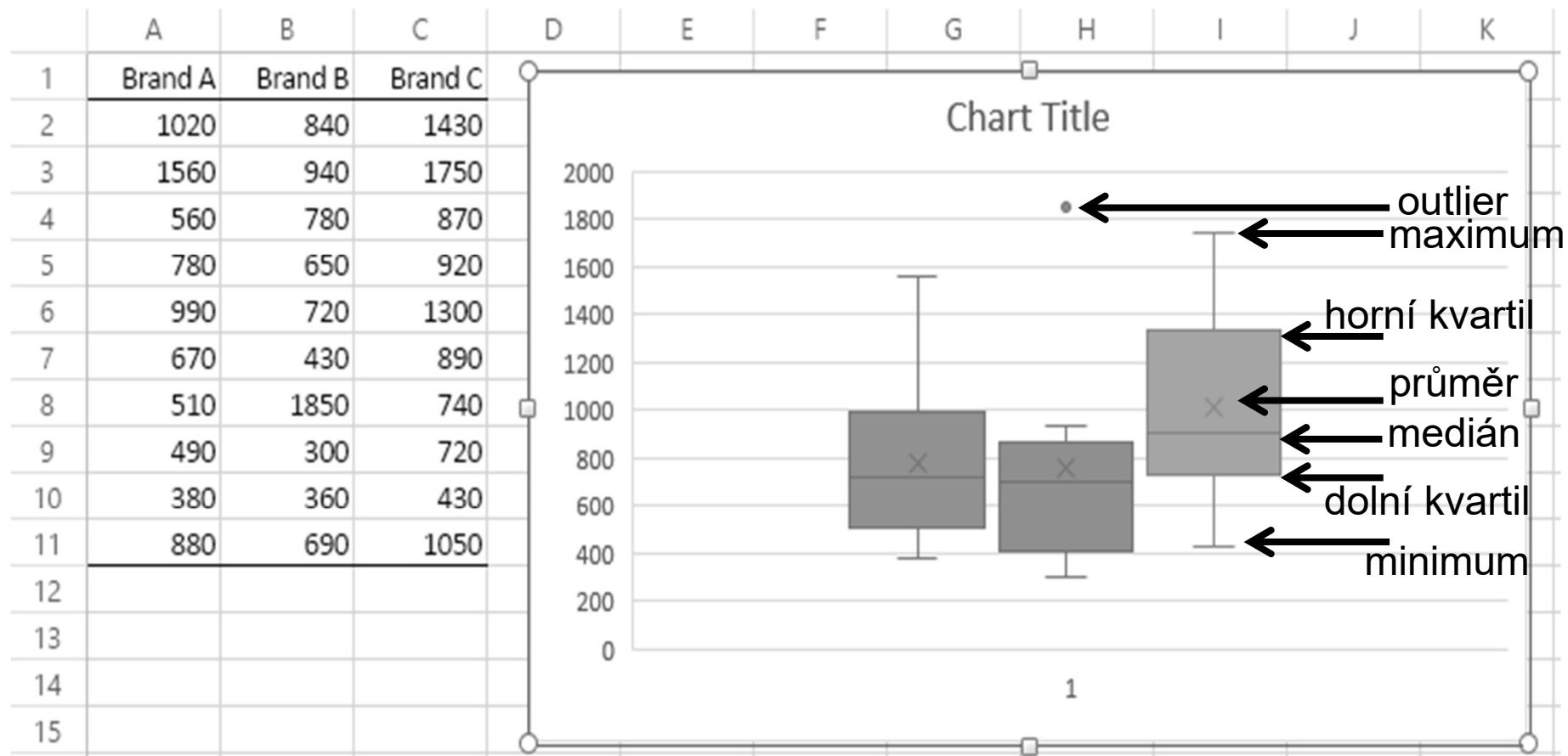
Charakteristiky číselných veličin

„KRABÍČKOVÝ GRAF“ (boxplot)

Excel 2016



Charakteristiky číselných veličin



Min, max = „neodlehle“ extrémy

Outlier = odlehlá hodnota

PRAVDĚPODOBNOST

Základní vlastnosti (JIŽ BYLO):

a) $P(\{ \}) = 0, \quad P(\Omega) = 1$

b) $0 \leq P(A) \leq 1$

c) $A \text{ je podmnožina } B \Rightarrow P(A) < P(B)$

d) $P(A') = 1 - P(A)$ (*doplňěk*)

e) $P(A \cap B) = P(A) \cdot P(B)$ (*A, B nezávislé*)

f) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

g) $P(A|B) = P(A \cap B) / P(B)$ (*podmínka*)

Podmíněná prst. (příklad)

Nechť v populaci je 60 % samců (M). Víme, že mezi samci je 10% albínů (A), zatímco mezi samicemi (M') je 20% albínů.

a) Určete, jaký je celkově podíl albínů.

b) Určete, jakou část z albínů tvoří samci.

Zadáno: $P(M)=0,6$ $P(M')=0,4$

$P(A|M)=0,1$ $P(A|M')=0,2$

a) $P(A \cap M)=0,1 \cdot 0,6=0,06$ $P(A \cap M')=0,2 \cdot 0,4=0,08$

$P(A)=0,06+0,08=\underline{0,14}$

b) $P(M|A)=P(A|M) \cdot P(M) / [P(A|M) \cdot P(M) + P(A|M') \cdot P(M')] =$
 $= P(A \cap M) / P(A) = 0,06 / 0,14 = \underline{0,43}$

Příklad 4 (rozklad) – JIŽ BYLO:

Ω : výsledky 2 hodů kostkou (X =počet šestek)

A_1 ...nehozena ani jedna šestka ($X=0$)

A_2 ...hozena jedna šestka ($X=1$)

A_3 ...hozeny dvě šestky ($X=2$)

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

$$P(A_1)=25/36 \quad \text{aneb} \quad P(X=0)=25/36$$

$$P(A_2)=10/36 \quad \text{aneb} \quad P(X=1)=10/36$$

$$P(A_3)=1/36 \quad \text{aneb} \quad P(X=2)=1/36$$

Výpočty pomocí základních vlastností:

- $P(A_3) = P(A \cap A) = P(A) \cdot P(A) = 1/6 \cdot 1/6 = 1/36$

...jde o průnik výskytu $A=\{6\}$ při 1.hodu a při 2.hodu,
tj. platí nezávislost

- $P(A_1) = P(A' \cap A') = P(A') \cdot P(A') = 5/6 \cdot 5/6 = 25/36$

...analog., ale v obou hodech má nastat doplněk

- $P(A_2) = P[(A \cap A') \cup (A' \cap A)] = P(A \cap A') + P(A' \cap A) =$
 $= P(A) \cdot P(A') + P(A') \cdot P(A) = 1/6 \cdot 5/6 + 5/6 \cdot 1/6 =$
 $= 5/36 + 5/36 = 10/36$

...buď hodíme šestku jen napoprvé, nebo jen
napodruhé

POKRAČUJEME:

$$P(X=0)=25/36$$

$$P(X=1)=10/36$$

$$P(X=2)=1/36$$

Zákon rozdělení: $\sum P_i=1$ (neb A_i tvoří rozklad Ω)

Modus=0 (nejpravděpodobnější hodnota X)

První obecný moment aneb: Střední hodnota:

$$EX = \sum x_i \cdot P_i, \quad i=1 \dots K$$

$$EX = 0 \cdot 25/36 + 1 \cdot 10/36 + 2 \cdot 1/36 = 12/36 = 1/3$$

Druhý obecný moment:

$$E(X^2) = \sum x_i^2 \cdot P_i, \quad i=1 \dots K$$

$$E(X^2) = 0^2 \cdot 25/36 + 1^2 \cdot 10/36 + 2^2 \cdot 1/36 = 14/36 = 7/18$$

Druhý centrovaný moment (rozptyl):

$$DX = \sum (x_i - EX)^2 \cdot P_i, \quad i=1 \dots K$$

$$\begin{aligned} DX &= (0 - 1/3)^2 \cdot 25/36 + \\ &\quad + (1 - 1/3)^2 \cdot 10/36 + \\ &\quad + (2 - 1/3)^2 \cdot 1/36 = \dots = 5/18 \end{aligned}$$

$$P(X=0)=25/36$$

$$P(X=1)=10/36$$

$$P(X=2)=1/36$$

Druhý centrovaný moment pomocí obecných:

$$DX = E(X^2) - (EX)^2$$

$$DX = 7/18 - (1/3)^2 = 5/18$$

Medián (obecně: kvantily)?

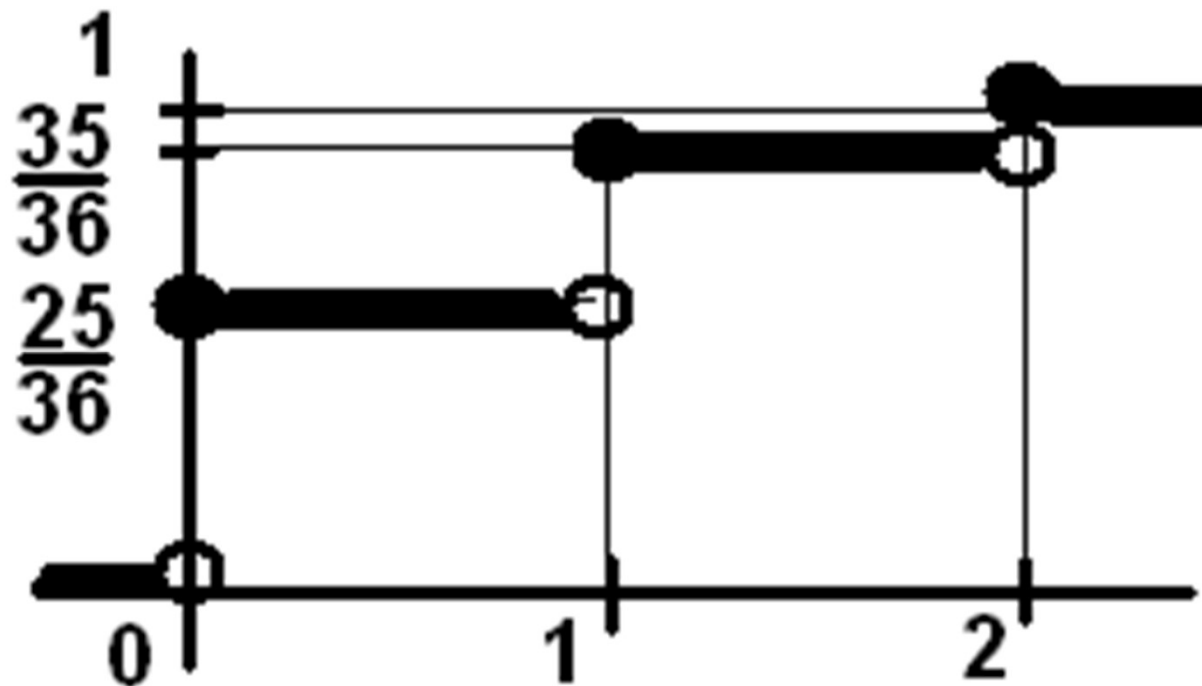
Distribuční funkce

$F(x) = P(X \leq x)$ = „kumulovaná pravděpodobnost“

$$F(0) = P(0) = 25/36$$

$$F(1) = P(0) + P(1) = 25/36 + 10/36 = 35/36$$

$$F(2) = P(0) + P(1) + P(2) = 25/36 + 10/36 + 1/36 = 1$$



Alternativní rozdělení

■ $X \sim \text{Alt}(\pi)$

A...sledovaný jev, $\pi = P(A)$

X =počet výskytů A při jediném pokusu

$X=0$ nebo 1

$$P(X=0) = P(A') = 1-\pi, \quad P(X=1) = \pi$$

$$EX = 0 \cdot (1-\pi) + 1 \cdot \pi = \pi,$$

$$E(X^2) = 0^2 \cdot (1-\pi) + 1^2 \cdot \pi = \pi,$$

$$DX = \pi - \pi^2 = \pi \cdot (1-\pi)$$

Binomické rozdělení

■ $X \sim \text{Bi}(n, \pi)$

$\pi = P(A)$, $A \dots$ sledovaný jev

$n =$ počet nezávislých pokusů (π se v nich nemění)

$X =$ počet výskytů A při n pokusech

$$X = 0, 1, \dots, n$$

$$P(X=x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$$

$$EX = n \cdot \pi$$

$$DX = n \cdot \pi \cdot (1-\pi)$$

Příklad: Počet šestek při 2 hodech kostkou. (=Př.4)

Příklad: Počet „úspěchů“ při losování s vracením.

Poznámka:

V případě, kdy vybíráme technikou bez vracení, je správný hypergeometrický model; pokud ale výběr probíhá z velké populace, použijeme i pak binomický model.

Proč? Když např. z 10.000 obyvatel, kde jevu A vyhovovalo 5.000 obyvatel (tj. $\pi=0,500$) ubereme 30 jedinců technikou bez vracení, bude poté jevu A vyhovovat podíl z rozmezí (dle postupu vybírání):

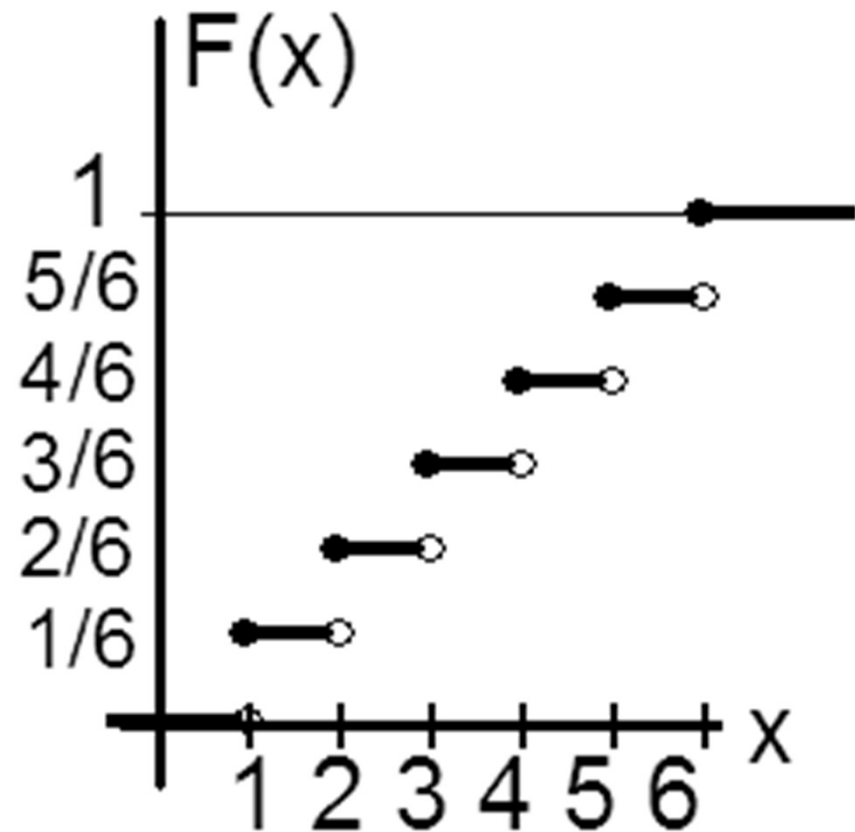
$$4.970 / 9.970 \quad \text{až} \quad 5.000 / 9.970$$

$$0,498 \quad \text{až} \quad 0,502$$

což se příliš nezměnilo oproti hodnotě 0,500 aneb pravděpodobnost výskytu jevu A téměř nezávisí na pořadí výběru (lze ji stále považovat za ono π)

Rovnoměrné rozdělení (kategoriální typ)

- $X \sim R(K)$
- $X = \text{hodnoty } 1, 2, \dots, K$
stejně pravděpodobné
- $P(X=x) = 1/K$,
 $EX = (1+K)/2$
- Příklad: $X = \text{výsledek}$
hodu kostkou ($K=6$).
Graf distribuční funkce
 $F(x) = P(X \leq x)$



Rovnoměrné rozdělení (spojitý typ)

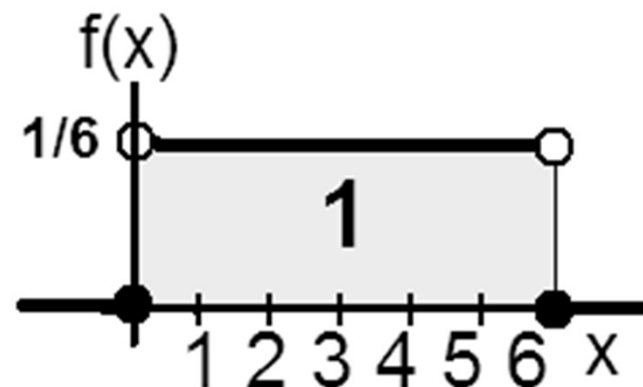
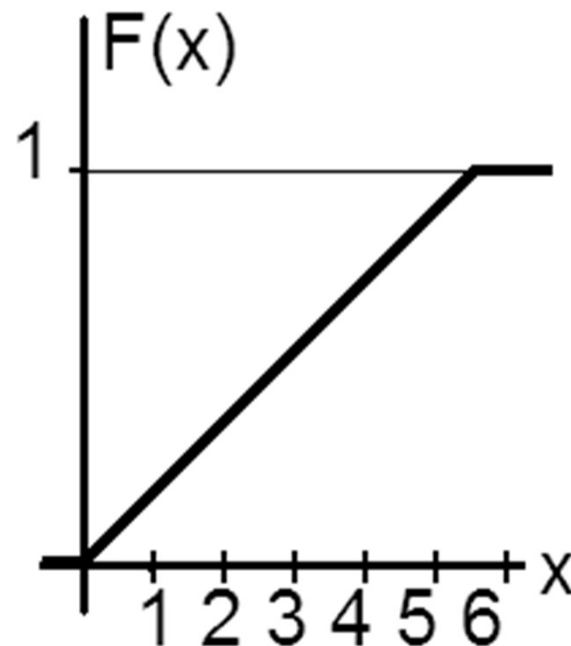
- $X \sim R(0, K)$
- X = kterékoli reálné číslo mezi 0 až K
- Distribuční funkce
$$F(x) = P(X \leq x) = x/K$$

(mezi 0 až K , jinde konstanta viz graf)

- Hustota

$$f(x) = F'(x) = 1/K$$

(mezi 0 až K , jinde 0)



Zákon rozdělení: $\int f(x)dx=1$... integrály určité
(zde meze integrálu: min X až max X)

Distribuční funkce:

$F(b) = P(X \leq b)$... všeobecně platná definice!

$F(b) = \int f(x) dx$... meze integrálu: min X až b

$F(b)$ = velikost plochy pod hustotou mezi minX až b

Důsledky pro určování pravděpodobností:

$P(X > a) = 1 - F(a) = \int f(x) dx$ (integrál od a do max X)

$P(a \leq X \leq b) = F(b) - F(a) = \int f(x) dx$ (integrál od a do b)

Střední hodnota: $EX = \int x \cdot f(x) dx$

Druhý obecný moment: $E(X^2) = \int x^2 \cdot f(x) dx$

Druhý centrovaný moment
(rozptyl): $DX = E(X^2) - (EX)^2$

Př. (rovnom.spojité veličina, meze: 0 až K)

$f(x) = 1/K$ pro všechna $x = 0$ až K ; $f(x) = 0$ jinde

$$EX = \int x \cdot 1/K dx = [x^2/(2K)] = K^2/(2K) - 0 = K/2$$

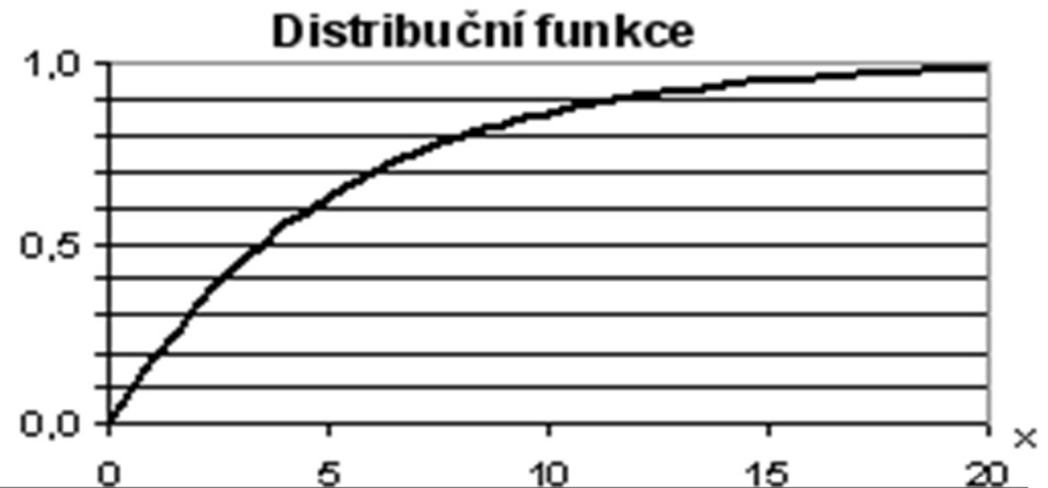
$$EX^2 = \int x^2 \cdot 1/K dx = [x^3/(3K)] = K^3/(3K) - 0 = K^2/3$$

$$DX = K^2/3 - (K/2)^2 = K^2/3 - K^2/4 = K^2/12$$

Exponenciální rozdělení

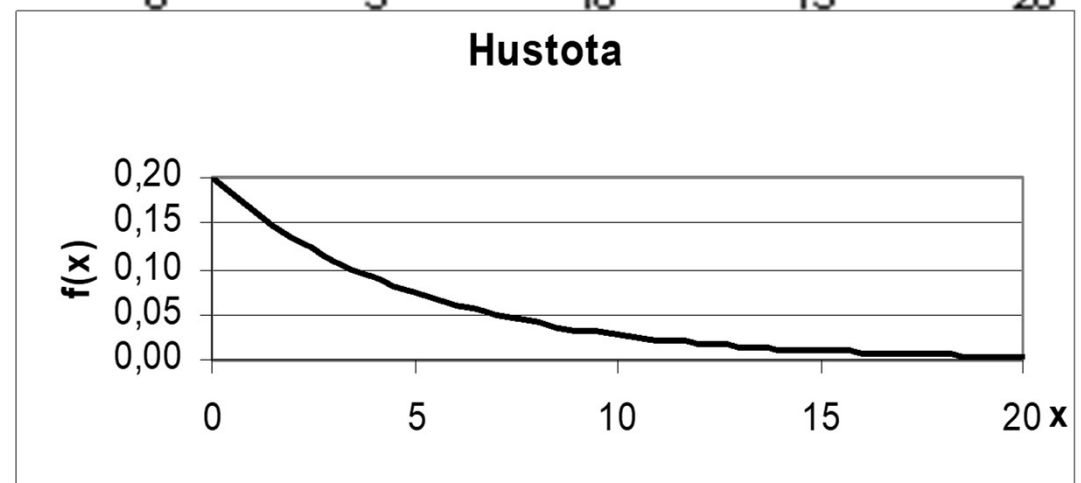
- $X \sim \text{Exp}(\delta)$
- X = jakákoli kladná hodnota (doba čekání)
- Distribuční funkce ($x > 0$)

$$F(x) = P(X \leq x) = 1 - e^{-x/\delta}$$



- Hustota ($x > 0$)

$$f(x) = F'(x) = e^{-x/\delta}/\delta$$



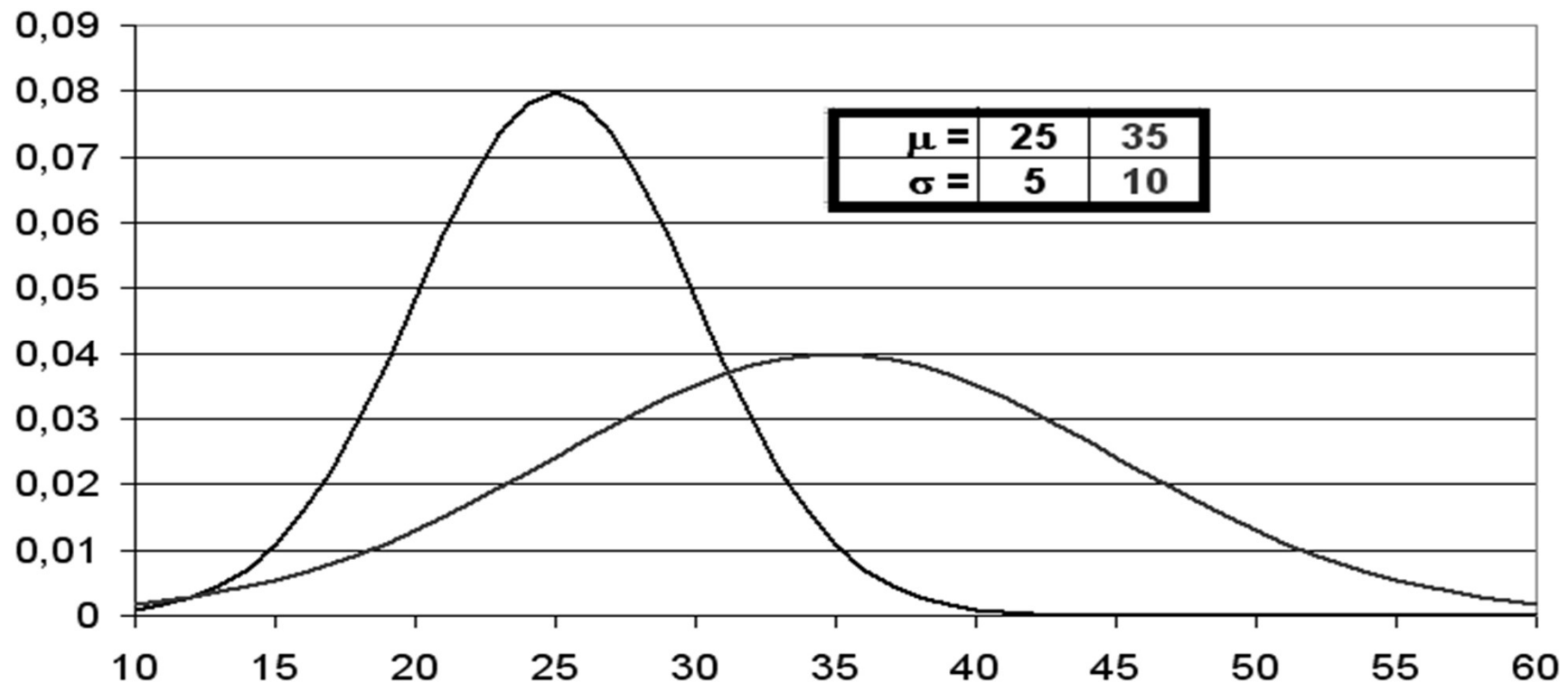
Gaussovo normální rozdělení

■ $X \sim N(\mu, \sigma^2)$

... $\mu = EX$, $\sigma^2 = DX$

Grafem hustoty $f(x)$ (vzorec viz skripta) Gaussova křivka

(centrována kolem $\mu = EX$, ale je to též modus – hodnota s max. hustotou a medián).



Gaussovo normální rozdělení

■ Příklad 5: Necht' je hmotnost $X \sim N(80, 100)$.

Určíme $P(X \leq 90)$, $P(X \leq 75)$ a $P(75 \leq X \leq 90)$.

Výpočty např. pomocí Excelu:

$$\begin{aligned} \text{a) } P(X \leq 90) &= F(90) = \text{NORM.DIST}(90; 80; 10; 1) = \\ &= 0,84 = 84 \% \end{aligned}$$

$$\begin{aligned} \text{b) } P(X \leq 75) &= F(75) = \text{NORM.DIST}(75; 80; 10; 1) = \\ &= 0,31 = 31 \% \end{aligned}$$

$$\begin{aligned} \text{c) } P(75 \leq X \leq 90) &= P(X \leq 90) - P(X \leq 75) = 0,84 - 0,31 = \\ &= 0,53 = 53 \%. \end{aligned}$$

Základy statistické indukce

- Pro stat.data hledáme vhodné pravděp. modely, odhadujeme hodnoty jejich parametrů či testujeme tvrzení o chování stat.veličin.
- Využíváme toho, že charakteristiky stat. dat (např. aritm.průměr) vykazují vlastnosti pravděpodobnostních rozdělení.

Základy statistické indukce

■ BODOVÉ ODHADY (tj. odhady jedním číslem)

T_n (např. aritm.průměr či medián) je z dat získaný bodový odhad pro neznámý parametr θ v pravděpodobnostním modelu pro sledovanou veličinu (např. pro střední hodnotu v normálním rozdělení).

Je to odhad nestranný $\Leftrightarrow E(T_n) = \theta$.

Základy statistické indukce

- Tabulka teoretických (neznámých a tudíž odhadovaných) parametrů a jejich nejvhodnějších (nestranných) odhadů:

PARAMETR θ	JEHO BODOVÝ ODHAD T_n
$\pi = P(A)$	$p =$ relativní četnost jevu A
μ (střední hodnota)	aritmetický průměr
σ^2 (rozptyl)	výběrový rozptyl: $s^2 = M_2 \cdot n / (n-1)$

Základy statistické indukce

■ Zákony velkých čísel

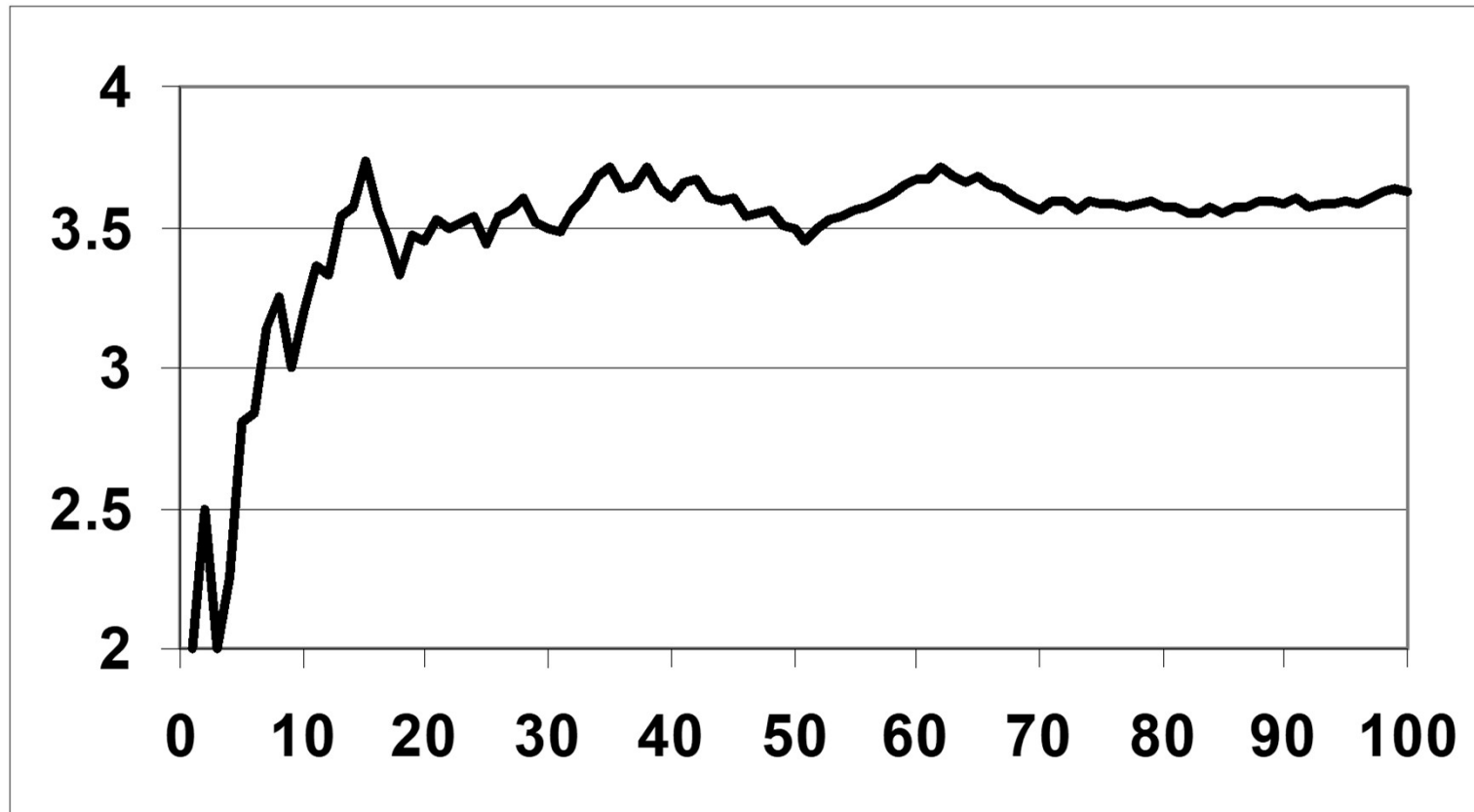
(chování bodových odhadů):

Např. průměr $\rightarrow$ střední hodnota (zde $EX=3,500$)

Pořadí hodu	1	2	3	4	5	...	98	99	100
Hozeno	2	3	1	3	5	...	5	5	3
Průměr	2,000	2,500	2,000	2,250	2,800	...	3,622	3,636	3,630

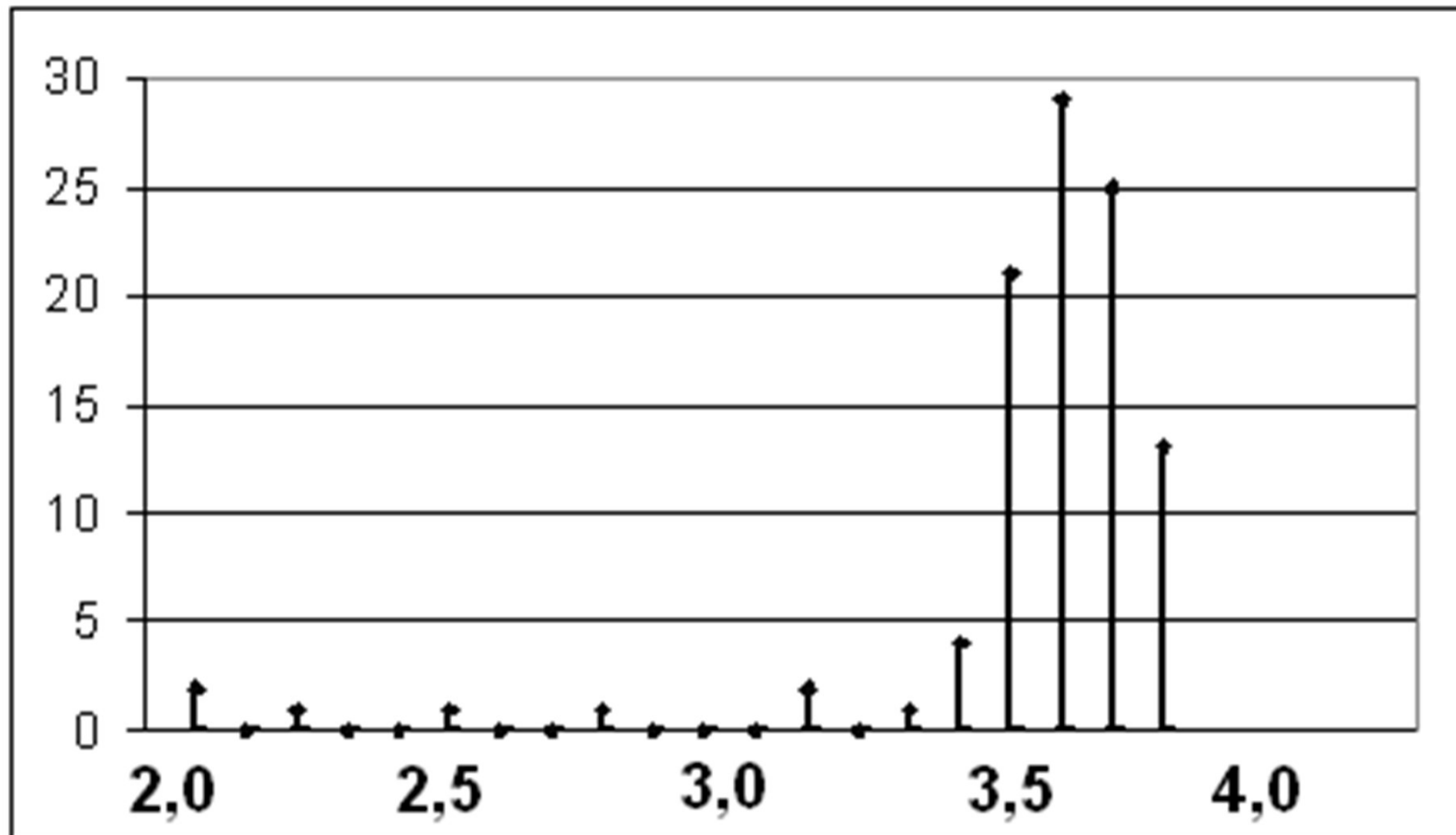
Základy statistické indukce

- Zákony velkých čísel (pokračování ilustrace):



Základy statistické indukce

- Zákony velkých čísel (pokr.-četnosti průměrů):



Základy statistické indukce

■ Centrální limitní věty (CLV)

Popisují asymptotické (tj. v limitě, v praxi pro „dostatečně velký“ počet stat. dat) chování testových charakteristik T_n jakožto náhodných veličin.

Např.

$$T_n = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})} \rightarrow U \sim N(0,1)$$