III. Část

ZPRACOVÁNÍ DAT ZÁKLADNÍMI STATISTICKÝMI METODAMI

Statistické zpracování dat: Číselné charakteristiky

Shrnují **vlastnosti** rozdělení pravděpodobnosti **do jednoho čísla**, které je snadno interpretovatelné a lze s ním pracovat jednodušeji než s funkčním vyjádřením.

Dvě nejznámější a nejpoužívanější charakteristiky, které odráží vlastnosti rozdělení pravděpodobnosti náhodné veličiny, jsou

- **střední hodnota** (mean value) a
- **rozptyl** (dispersion, variance).

Dalšími charakteristikami jsou **koeficienty**

- **šikmosti** (skewness) a
- **špičatosti** (kurtosis).

V neposlední řadě se i mezi laickou veřejností stále více využívají **kvantily**, speciálně

- **kvartily**,
- **decily** a
- **percentily**.

Nakonec uvedeme **modus**, což je nejpravděpodobnější hodnota, kterou lze v daném souboru očekávat.

Statistické zpracování dat: Číselné charakteristiky

Shrnují **vlastnosti** rozdělení pravděpodobnosti **do jednoho čísla**, které je snadno interpretovatelné a lze s ním pracovat jednodušeji než s funkčním vyjádřením.

Dvě nejznámější a nejpoužívanější charakteristiky, které odráží vlastnosti rozdělení pravděpodobnosti náhodné veličiny, jsou

- **střední hodnota** (mean value) a
- **rozptyl** (dispersion, variance).

Dalšími charakteristikami jsou **koeficienty**

- **šikmosti** (skewness) a
- **špičatosti** (kurtosis).

momentové charakteristiky

V neposlední řadě se i mezi laickou veřejností stále více využívají **kvantily**, speciálně

- **kvartily**,
- **decily** a
- **percentily**.

Nakonec uvedeme **modus**, což je nejpravděpodobnější hodnota, kterou lze v daném souboru očekávat.

Statistické zpracování dat: Číselné charakteristiky

Shrnují **vlastnosti** rozdělení pravděpodobnosti **do jednoho čísla**, které je snadno interpretovatelné a lze s ním pracovat jednodušeji než s funkčním vyjádřením.

Dvě nejznámější a nejpoužívanější charakteristiky, které odráží vlastnosti rozdělení pravděpodobnosti náhodné veličiny, jsou

- **střední hodnota** (mean value) a
- **rozptyl** (dispersion, variance).

Dalšími charakteristikami jsou **koeficienty**

- **šikmosti** (skewness) a
- **špičatosti** (kurtosis).

momentové charakteristiky

V neposlední řadě se i mezi laickou veřejností stále více využívají **kvantily**, speciálně

- **kvartily**,
- **decily** a
- **percentily**.

Nakonec uvedeme **modus**, což je nejpravděpodobnější hodnota, kterou lze v daném souboru očekávat.

Statistické zpracování dat: **Číselné charakteristiky**

Střední hodnota = průměr
První obecný moment

$$\bar{x} = E(X) = \frac{\sum_i x_i}{n}$$

$$= \sum_i x_i \cdot p(x_i)$$

Určuje průměrnou hodnotu dat.
Je lehce ovlivnitelný extrémními hodnotami.

Statistické zpracování dat: Číselné charakteristiky

Střední hodnota = průměr
První obecný moment

$$\bar{x} = E(X) = \frac{\sum_i x_i}{n}$$

$$= \sum_i x_i \cdot p(x_i)$$

Určuje průměrnou hodnotu dat.
Je lehce ovlivnitelný extrémními hodnotami.

Rozptyl
Druhý **centrální** moment

$$\mathrm{D}(X) = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

$$= \sum_i (x_i - \bar{x})^2 \cdot p(x_i)$$

Popisuje rozptýlenost nejčastějších hodnot okolo průměru.
POZOR! Jednotky jsou umocněné na druhou.

Statistické zpracování dat: # Číselné charakteristiky

Střední hodnota = průměr
První obecný moment

$$\bar{x} = E(X) = \frac{\sum_i x_i}{n}$$

$$= \sum_i x_i \cdot p(x_i)$$

Určuje průměrnou hodnotu dat.
Je lehce ovlivnitelný extrémními hodnotami.

Rozptyl
Druhý **centrální** moment

$$D(X) = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

$$= \sum_i (x_i - \bar{x})^2 \cdot p(x_i)$$

Popisuje rozptýlenost nejčastějších hodnot okolo průměru.
POZOR! Jednotky jsou umocněné na druhou.

Statistické zpracování dat: Číselné charakteristiky

Střední hodnota = průměr
První obecný moment

$$\bar{x} = E(X) = \frac{\sum_i x_i}{n}$$

$$= \sum_i x_i \cdot p(x_i)$$

Určuje průměrnou hodnotu dat.
Je lehce ovlivnitelný extrémními hodnotami.

Rozptyl
Druhý centrální moment

$$D(X) = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

$$= \sum_i (x_i - \bar{x})^2 \cdot p(x_i)$$

Popisuje rozptýlenost nejčastějších hodnot okolo průměru.
POZOR! Jednotky jsou umocněné na druhou.

Směrodatná odchylka

$$\sigma = \sqrt{D(X)}$$

Popisuje rozptýlenost nejčastějších hodnot okolo průměru.

Statistické zpracování dat: Číselné charakteristiky

Střední hodnota = průměr

První obecný moment

$$\bar{x} = E(X) = \frac{\sum_i x_i}{n}$$

$$= \sum_i x_i \cdot p(x_i)$$

Určuje průměrnou hodnotu dat.

Je lehce ovlivnitelný extrémními hodnotami.

Rozptyl

Druhý **centrální** moment

$$\mathrm{D}(X) = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

$$= \sum_i (x_i - \bar{x})^2 \cdot p(x_i)$$

Popisuje rozptýlenost nejčastějších hodnot okolo průměru.

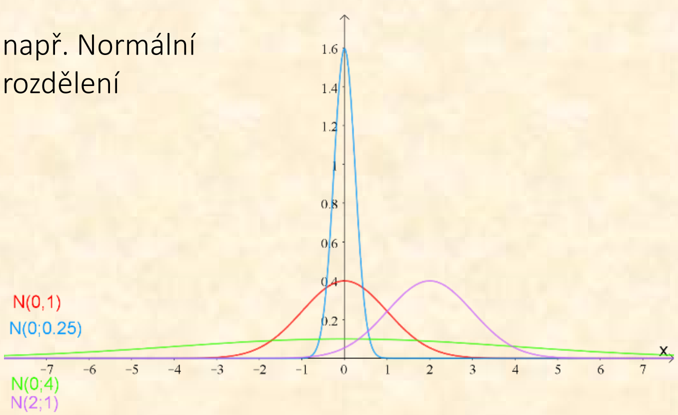
POZOR! Jednotky jsou umocněné na druhou.

Směrodatná odchylka

$$\sigma = \sqrt{\mathrm{D}(X)}$$

Popisuje rozptýlenost nejčastějších hodnot okolo průměru.

např. Normální rozdělení



Statistické zpracování dat: Číselné charakteristiky

Shrnují **vlastnosti** rozdělení pravděpodobnosti **do jednoho čísla**, které je snadno interpretovatelné a lze s ním pracovat jednodušeji než s funkčním vyjádřením.

Dvě nejznámější a nejpoužívanější charakteristiky, které odráží vlastnosti rozdělení pravděpodobnosti náhodné veličiny, jsou

- **střední hodnota** (mean value) a
- **rozptyl** (dispersion, variance).

Dalšími charakteristikami jsou **koeficienty**

- **šikmosti** (skewness) a
- **špičatosti** (kurtosis).

momentové charakteristiky

V neposlední řadě se i mezi laickou veřejností stále více využívají **kvantily**, speciálně

- **kvartily**,
- **decily** a
- **percentily**.

Nakonec uvedeme **modus**, což je nejpravděpodobnější hodnota, kterou lze v daném souboru očekávat.

Statistické zpracování dat: Číselné charakteristiky

Shrnují **vlastnosti** rozdělení pravděpodobnosti **do jednoho čísla**, které je snadno interpretovatelné a lze s ním pracovat jednodušeji než s funkčním vyjádřením.

Dvě nejznámější a nejpoužívanější charakteristiky, které odráží vlastnosti rozdělení pravděpodobnosti náhodné veličiny, jsou

- **střední hodnota** (mean value) a
- **rozptyl** (dispersion, variance).

Dalšími charakteristikami jsou **koeficienty**

- **šikmosti** (skewness) a
- **špičatosti** (kurtosis).

momentové charakteristiky

V neposlední řadě se i mezi laickou veřejností stále více využívají **kvantily**, speciálně

- **kvartily**,
- **decily** a
- **percentily**.

Nakonec uvedeme **modus**, což je nejpravděpodobnější hodnota, kterou lze v daném souboru očekávat.

Koeficient šikmosti

Normovaný třetí **centrální** moment

$$A = \frac{1}{\sigma^3}\frac{\sum_i (x_i - \overline{x})^3}{n} = \frac{1}{\sigma^3}\sum_i (x_i - \overline{x})^3 \cdot p(x_i)$$

Popisuje míru a směr zešikmení pravděpodobnostní křivky. Zešikmení

- směrem k maximálním hodnotám (**doprava**) pro $\boldsymbol{A < 0}$
- směrem k minimálním hodnotám (**vlevo**) pro $\boldsymbol{A > 0}$
- **symetrii** pro $\boldsymbol{A = 0}$

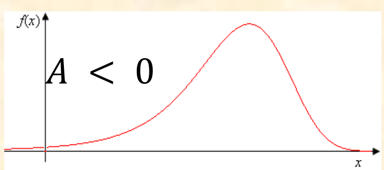
Koeficient šikmosti

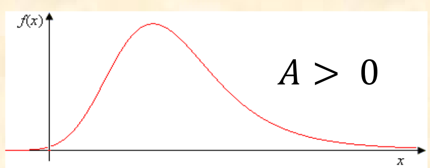
Normovaný třetí centrální moment

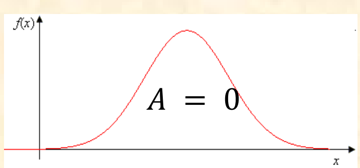
$$A = \frac{1}{\sigma^3}\frac{\sum_i (x_i - \overline{x})^3}{n}$$
$$= \frac{1}{\sigma^3}\sum_i (x_i - \overline{x})^3 \cdot p(x_i)$$

Popisuje míru a směr zešikmení pravděpodobnostní křivky. Zešikmení

- směrem k maximálním hodnotám (**doprava**) pro $\boldsymbol{A < 0}$
- směrem k minimálním hodnotám (**vlevo**) pro $\boldsymbol{A > 0}$
- **symetrii** pro $\boldsymbol{A = 0}$







Statistické zpracování dat: # Číselné charakteristiky

Koeficient šikmosti

Normovaný třetí **centrální** moment

$$A=\frac{1}{\sigma^3}\frac{\sum_i (x_i-\bar{x})^3}{n}$$

$$=\frac{1}{\sigma^3}\sum_i (x_i-\bar{x})^3\cdot p(x_i)$$

Popisuje míru a směr zešikmení pravděpodobnostní křivky. Zešikmení

- směrem k maximálním hodnotám (**doprava**) pro $\boldsymbol{A<0}$
- směrem k minimálním hodnotám (**vlevo**) pro $\boldsymbol{A>0}$
- **symetrii** pro $\boldsymbol{A=0}$

Koeficient špičatosti

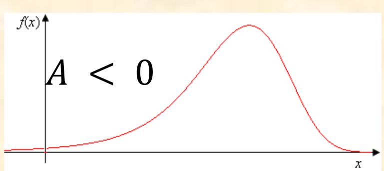
Normovaný čtvrtý **centrální** moment

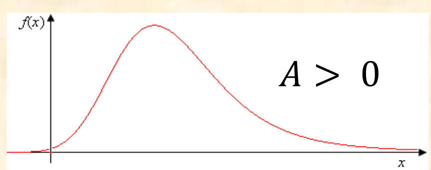
$$\bar{e}=\frac{1}{\sigma^4}\frac{\sum_i (x_i-\bar{x})^4}{n}-3$$

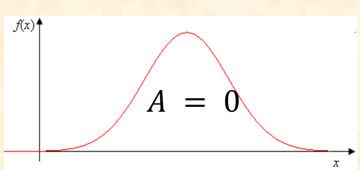
$$=\frac{1}{\sigma^4}\sum_i (x_i-\bar{x})^4\cdot p(x_i)-3$$

Popisuje míru špičatosti pravděpodobnostní křivky. Podle výsledku

- $\bar{e}>0$ špičatější,
- $\bar{e}<0$ placatější,
- $\bar{e}=0$ odpovídá $N(0{,}1)$.







Číselné charakteristiky

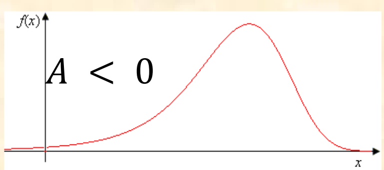
Koeficient šikmosti

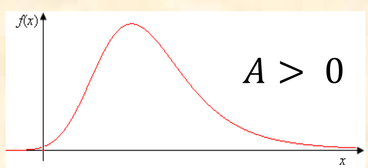
Normovaný třetí centrální moment

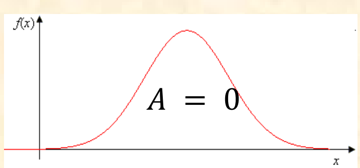
$$A = \frac{1}{\sigma^3} \frac{\sum_i (x_i - \overline{x})^3}{n} = \frac{1}{\sigma^3} \sum_i (x_i - \overline{x})^3 \cdot p(x_i)$$

Popisuje míru a směr zešikmení pravděpodobnostní křivky. Zešikmení

- směrem k maximálním hodnotám (**doprava**) pro $\boldsymbol{A < 0}$
- směrem k minimálním hodnotám (**vlevo**) pro $\boldsymbol{A > 0}$
- **symetrii** pro $\boldsymbol{A = 0}$







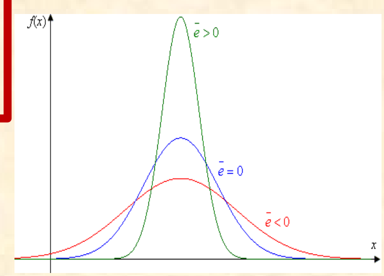
Koeficient špičatosti

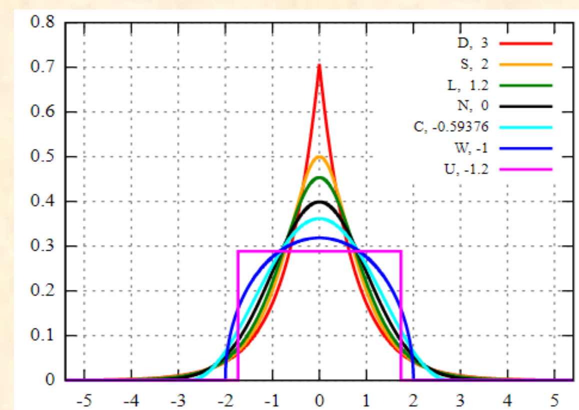
Normovaný čtvrtý centrální moment

$$\bar{e} = \frac{1}{\sigma^4} \frac{\sum_i (x_i - \overline{x})^4}{n} - 3 = \frac{1}{\sigma^4} \sum_i (x_i - \overline{x})^4 \cdot p(x_i) - 3$$

Popisuje míru špičatosti pravděpodobnostní křivky. Podle výsledku

- $\bar{e} > 0$ špičatější,
- $\bar{e} < 0$ placatější,
- $\bar{e} = 0$ odpovídá $N(0,1)$.





Statistické zpracování dat: Číselné charakteristiky

Shrnují **vlastnosti** rozdělení pravděpodobnosti **do jednoho čísla**, které je snadno interpretovatelné a lze s ním pracovat jednodušeji než s funkčním vyjádřením.

Dvě nejznámější a nejpoužívanější charakteristiky, které odráží vlastnosti rozdělení pravděpodobnosti náhodné veličiny, jsou

- **střední hodnota** (mean value) a
- **rozptyl** (dispersion, variance).

Dalšími charakteristikami jsou **koeficienty**

- **šikmosti** (skewness) a
- **špičatosti** (kurtosis).

momentové charakteristiky

V neposlední řadě se i mezi laickou veřejností stále více využívají **kvantily**, speciálně

- **kvartily**,
- **decily** a
- **percentily**.

Nakonec uvedeme **modus**, což je nejpravděpodobnější hodnota, kterou lze v daném souboru očekávat.

Statistické zpracování dat: Číselné charakteristiky

Shrnují **vlastnosti** rozdělení pravděpodobnosti **do jednoho čísla**, které je snadno interpretovatelné a lze s ním pracovat jednodušeji než s funkčním vyjádřením.

Dvě nejznámější a nejpoužívanější charakteristiky, které odráží vlastnosti rozdělení pravděpodobnosti náhodné veličiny, jsou

- **střední hodnota** (mean value) a
- **rozptyl** (dispersion, variance).

Dalšími charakteristikami jsou **koeficienty**

- **šikmosti** (skewness) a
- **špičatosti** (kurtosis).

momentové charakteristiky

V neposlední řadě se i mezi laickou veřejností stále více využívají **kvantily**, speciálně

- **kvartily**,
- **decily** a
- **percentily**.

Nakonec uvedeme **modus**, což je nejpravděpodobnější hodnota, kterou lze v daném souboru očekávat.

Číselné charakteristiky

p-Kvantil

hodnota x_p z x-ové osy, pro kterou je

$$F(x_p) = p$$

Určuje hranici, jejíž hodnotu nebo menší má $100p$ procent čísel.

$$F(x_p) = P(X \leq x_p) = p$$

Statistické zpracování dat: Číselné charakteristiky

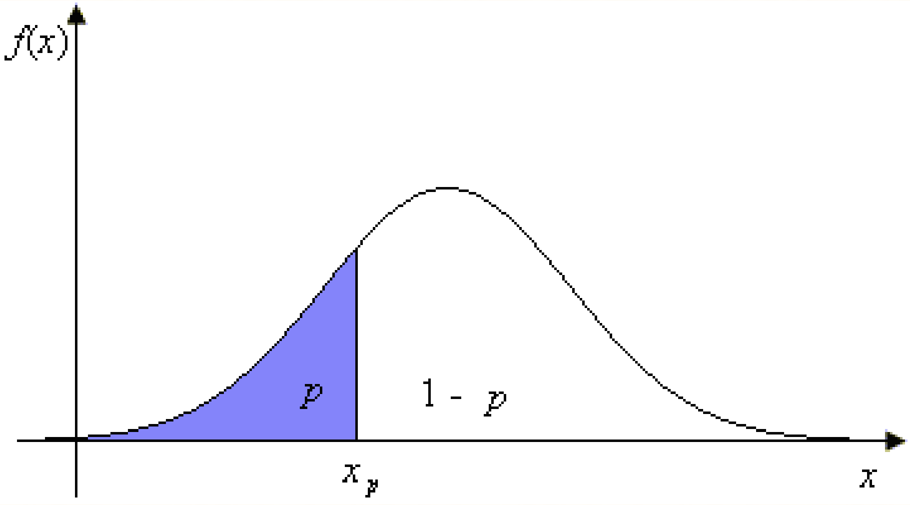
p-Kvantil

hodnota x_p z x-ové osy, pro kterou je

$$F(x_p) = p$$

Určuje hranici, jejíž hodnotu nebo menší má $100p$ procent čísel.

$$F(x_p) = P(X \leq x_p) = p$$



Statistické zpracování dat: Číselné charakteristiky

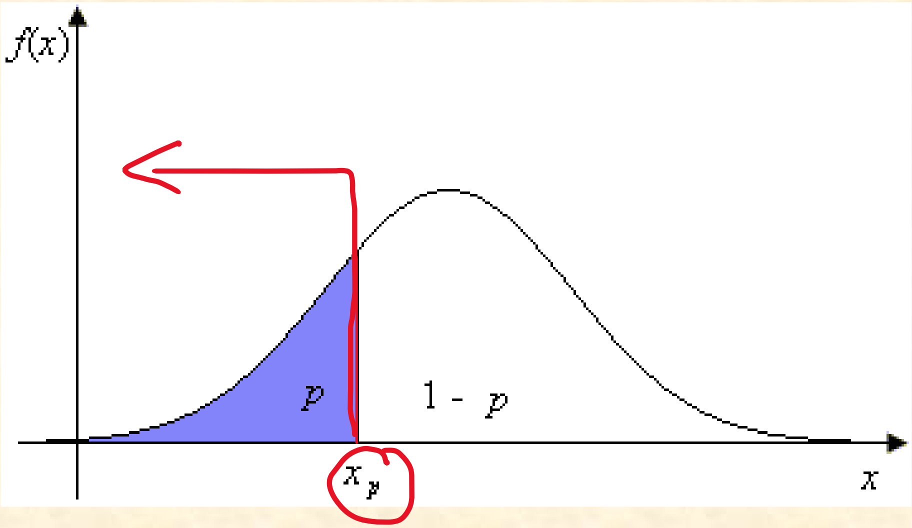
p-Kvantil

hodnota x_p z x-ové osy, pro kterou je

$$F(x_p) = p$$

Určuje hranici, jejíž hodnotu nebo menší má $100p$ procent čísel.

$$F(x_p) = P(X \leq x_p) = p$$



Číselné charakteristiky

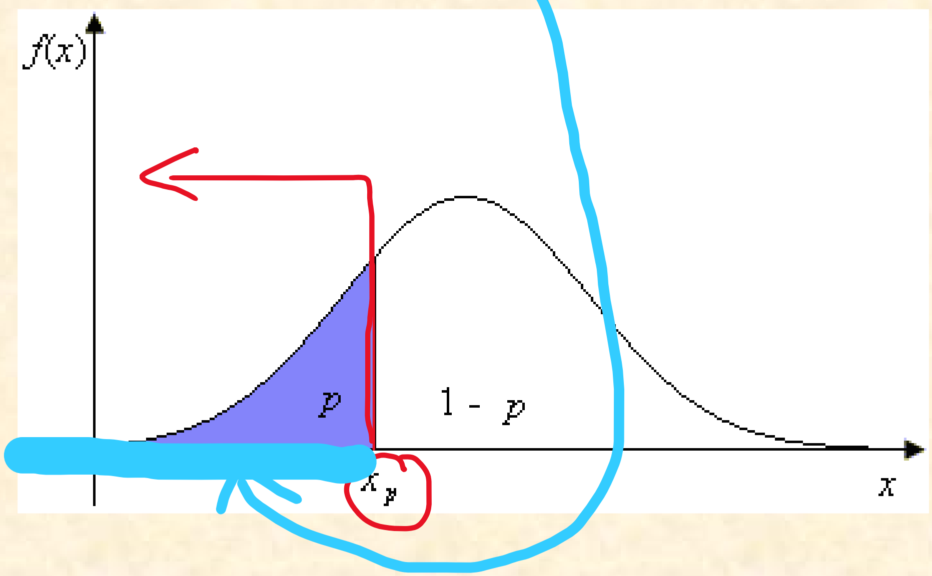
p-Kvantil

hodnota x_p z x-ové osy, pro kterou je

$$F(x_p) = p$$

Určuje hranici, jejíž hodnotu nebo menší má $100p$ procent čísel.

$$F(x_p) = P(X \leq x_p) = p$$



Statistické zpracování dat: Číselné charakteristiky

Shrnují **vlastnosti** rozdělení pravděpodobnosti **do jednoho čísla**, které je snadno interpretovatelné a lze s ním pracovat jednodušeji než s funkčním vyjádřením.

Dvě nejznámější a nejpoužívanější charakteristiky, které odráží vlastnosti rozdělení pravděpodobnosti náhodné veličiny, jsou

- **střední hodnota** (mean value) a
- **rozptyl** (dispersion, variance).

Dalšími charakteristikami jsou **koeficienty**

- **šikmosti** (skewness) a
- **špičatosti** (kurtosis).

momentové charakteristiky

V neposlední řadě se i mezi laickou veřejností stále více využívají **kvantily**, speciálně

- **kvartily**,
- **decily** a
- **percentily**.

Nakonec uvedeme **modus**, což je nejpravděpodobnější hodnota, kterou lze v daném souboru očekávat.

Statistické zpracování dat: Číselné charakteristiky

Shrnují **vlastnosti** rozdělení pravděpodobnosti **do jednoho čísla**, které je snadno interpretovatelné a lze s ním pracovat jednodušeji než s funkčním vyjádřením.

Dvě nejznámější a nejpoužívanější charakteristiky, které odráží vlastnosti rozdělení pravděpodobnosti náhodné veličiny, jsou

- **střední hodnota** (mean value) a
- **rozptyl** (dispersion, variance).

Dalšími charakteristikami jsou **koeficienty**

- **šikmosti** (skewness) a
- **špičatosti** (kurtosis).

momentové charakteristiky

V neposlední řadě se i mezi laickou veřejností stále více využívají **kvantily**, speciálně

- **kvartily**,
- **decily** a
- **percentily**.

Nakonec uvedeme **modus**, což je nejpravděpodobnější hodnota, kterou lze v daném souboru očekávat.

Modus

hodnota z x-ové osy, pro kterou je hodnota pravděpodobnostní funkce maximální

$$\mathrm{p}(Mo) = max$$

Jinými slovy jde o nejčastěji se vyskytující hodnotu.

Data mohou být

- uni-modální
- bi-modální
- … atd

Statistické zpracování dat: Číselné charakteristiky

Příklad 1: Pravděpodobnost zásahu cíle při každém ze čtyř výstřelů je 0,8. Označme náhodnou veličinou *X* označme počet zásahů cíle.

a) Určete rozdělení pravděpodobností náhodné veličiny.

b) Vypočtěte její střední hodnotu, rozptyl a směrodatnou odchylku.

Příklad 2: V městě byl po dobu 60ti dnů evidován počet dopravních. Podle počtu nehod v jednom dni byla sestavena následující tabulka:

počet nehod za den	0	1	2	3	4	5	6
počet dnů s uvedeným počtem nehod	4	28	10	7	6	4	1

Sestavte pravděpodobnostní funkci této náhodné veličiny a vypočtěte střední hodnotu a směrodatnou odchylku.

Příklad 3: (opakování z minulé hodiny) Náhodná veličina X má rozdělení N(0, 1). Určete:
a) $P(X < 2{,}31)$

b) $P(X < -1{,}1)$

c) $P(-0{,}41 < X < 2{,}92)$

Příklad 4: Trolejbusy odjíždějí ze zastávky v 10 minutových intervalech. Cestující může na zastávku přijít v libovolném okamžiku. Určete E(x) a D(x) doby čekání na odjezd trolejbusu.

Statistické zpracování dat: Číselné charakteristiky

Příklad 1: Pravděpodobnost zásahu cíle při každém ze čtyř výstřelů je 0,8. Označme náhodnou veličinou *X* označme počet zásahů cíle.

X……. počet zásahů ze 4 pokusů

a) Určete rozdělení pravděpodobností náhodné veličiny $X \sim Bi(4\,;0{,}8\,)$

b) Vypočtěte její střední hodnotu, rozptyl a směrodatnou odchylku.

Střední hodnota $E(X) = 3{,}2$ zásahů směrodatná odchylka $\sigma = 0{,}8$ zásahů

Příklad 2: V městě byl po dobu 60ti dnů evidován počet dopravních. Podle počtu nehod v jednom dni byla sestavena následující tabulka:

počet nehod za den	0	1	2	3	4	5	6
počet dnů s uvedeným počtem nehod	4	28	10	7	6	4	1

Sestavte pravděpodobnostní funkci této náhodné veličiny a vypočtěte střední hodnotu a směrodatnou odchylku. Vizte excelovský soubor.

Příklad 3: (opakování z minulé hodiny) Náhodná veličina X má rozdělení N(0, 1). Určete:

a) P(X < 2,31) 99%

b) P(X < -1,1) 13,5%

c) P(-0,41 < X < 2,92) 65,7%

Příklad 4: Trolejbusy odjíždějí ze zastávky v 10 minutových intervalech. Cestující může na zastávku přijít v libovolném okamžiku. Určete E(x) a D(x) doby čekání na odjezd trolejbusu.

Vizte excelovský soubor.

Statistické zpracování dat: Zpracování statistického souboru

2 způsoby zadání hodnot
statistického souboru
(čísel majících stejný význam, která jsou určena ke zpracování)

Statistické zpracování dat: Zpracování statistického souboru

2 způsoby zadání hodnot
statistického souboru
(čísel majících stejný význam, která jsou určena ke zpracování)

hodnoty
statistického znaku
jsou uvedeny tolikrát, kolikrát byly pozorovány

každá hodnota
statistického znaku
je uvedena jen jednou a k ní je vždy uvedeno, kolikrát byla pozorována tzv. četnost pozorování

Statistické zpracování dat: Zpracování statistického souboru

2 způsoby zadání hodnot
statistického souboru
(čísel majících stejný význam, která jsou určena ke zpracování)

hodnoty
statistického znaku
jsou uvedeny tolikrát, kolikrát byly pozorovány

počet prodaných baget typu BIG za den
12
15
23
12
16
15
15
14
8
8
12

každá hodnota
statistického znaku
je uvedena jen jednou a k ní je vždy uvedeno, kolikrát byla pozorována tzv. četnost pozorování

Statistické zpracování dat: # Zpracování statistického souboru

2 způsoby zadání hodnot **statistického souboru** (čísel majících stejný význam, která jsou určena ke zpracování)

hodnoty **statistického znaku** jsou uvedeny tolikrát, kolikrát byly pozorovány

počet prodaných baget typu BIG za den
12
15
23
12
16
15
15
14
8
8
12

každá hodnota **statistického znaku** je uvedena jen jednou a k ní je vždy uvedeno, kolikrát byla pozorována tzv. četnost pozorování

počet prodaných baget typu BIG za den	
počet baget	počet dní
12	3
15	3
23	1
16	1
14	1
8	2

Statistické zpracování dat: Zpracování statistického souboru

2 způsoby zadání hodnot **statistického souboru** (čísel majících stejný význam, která jsou určena ke zpracování)

hodnoty **statistického znaku** jsou uvedeny tolikrát, kolikrát byly pozorovány

počet prodaných baget typu BIG za den
12
15
23
12
16
15
15
14
8
8
12

seřadit podle 1. sloupce

každá hodnota **statistického znaku** je uvedena jen jednou a k ní je vždy uvedeno, kolikrát byla pozorována tzv. četnost pozorování

počet prodaných baget typu BIG za den	
počet baget	počet dní
12	3
15	3
23	1
16	1
14	1
8	2

Statistické zpracování dat: # Zpracování statistického souboru

2 způsoby zadání hodnot
statistického souboru
(čísel majících stejný význam, která jsou určena ke zpracování)

hodnoty
statistického znaku
jsou uvedeny tolikrát, kolikrát byly pozorovány

počet prodaných baget typu BIG za den
12
15
23
12
16
15
15
14
8
8
12

každá hodnota
statistického znaku
je uvedena jen jednou a k ní je vždy uvedeno, kolikrát byla pozorována tzv. četnost pozorování

počet prodaných baget typu BIG za den	
počet baget	počet dní
8	2
12	3
14	1
15	3
16	1
23	1

Statistické zpracování dat:
Zpracování statistického souboru

2 způsoby zadání hodnot **statistického souboru** (čísel majících stejný význam, která jsou určena ke zpracování)

hodnoty **statistického znaku** jsou uvedeny tolikrát, kolikrát byly pozorovány

počet prodaných baget typu BIG za den
12
15
23
12
16
15
15
14
8
8
12

statistický soubor

každá hodnota **statistického znaku** je uvedena jen jednou a k ní je vždy uvedeno, kolikrát byla pozorována tzv. četnost pozorování

počet prodaných baget typu BIG za den	
počet baget	počet dní
8	2
12	3
14	1
15	3
16	1
23	1

Zpracování statistického souboru

2 způsoby zadání hodnot **statistického souboru** (čísel majících stejný význam, která jsou určena ke zpracování)

hodnoty statistického znaku jsou uvedeny tolikrát, kolikrát byly pozorovány

x počet prodaných baget typu BIG za den
12
15
23
12
16
15
15
14
8
8
12

statistický soubor

statistický znak x

každá hodnota statistického znaku je uvedena jen jednou a k ní je vždy uvedeno, kolikrát byla pozorována tzv. četnost pozorování

počet prodaných baget typu BIG za den	
x počet baget	počet dní
8	2
12	3
14	1
15	3
16	1
23	1

Statistické zpracování dat: Zpracování statistického souboru

2 způsoby zadání hodnot
statistického souboru
(čísel majících stejný význam, která jsou určena ke zpracování)

hodnoty statistického znaku jsou uvedeny tolikrát, kolikrát byly pozorovány

každá hodnota statistického znaku je uvedena jen jednou a k ní je vždy uvedeno, kolikrát byla pozorována tzv. četnost pozorování

x počet prodaných baget typu BIG za den
12
15
23
12
16
15
15
14
8
8
12

x_i

statistický soubor

statistický znak x

hodnoty statistického znaku x_i

počet prodaných baget typu BIG za den	
x počet baget	počet dní
x_1 8	2
x_2 12	3
x_3 14	1
x_4 15	3
x_5 16	1
x_6 23	1

Zpracování statistického souboru

2 způsoby zadání hodnot **statistického souboru** (čísel majících stejný význam, která jsou určena ke zpracování)

hodnoty statistického znaku jsou uvedeny tolikrát, kolikrát byly pozorovány

každá hodnota statistického znaku je uvedena jen jednou a k ní je vždy uvedeno, kolikrát byla pozorována tzv. četnost pozorování

x počet prodaných baget typu BIG za den
12
15
23
12
16
15
15
14
8
8
12

x_i

statistický soubor

statistický znak x

hodnoty statistického znaku x_i

	počet prodaných baget typu BIG za den		
	x počet baget	počet dní	
x_1	8	2	f_1
x_2	12	3	f_2
x_3	14	1	f_3
x_4	15	3	f_4
x_5	16	1	f_5
x_6	23	1	f_6

četnosti statistického znaku f_i

Zpracování statistického souboru

2 způsoby zadání hodnot **statistického souboru** (čísel majících stejný význam, která jsou určena ke zpracování)

hodnoty **statistického znaku** jsou uvedeny tolikrát, kolikrát byly pozorovány

každá hodnota … je … a

n ….. počet různých hodnot statistického znaku

$N = \sum f_i$ ….. rozsah souboru (počet provedených měření)

x počet prodaných baget typu BIG za den
12
15
23
12
16
15
15
14
8
8
12

x_i

statistický soubor

statistický znak x

hodnoty statistického znaku x_i

	počet prodaných baget typu BIG za den		
	x počet baget	počet dní	
x_1	8	2	f_1
x_2	12	3	f_2
x_3	14	1	f_3
x_4	15	3	f_4
x_5	16	1	f_5
x_6	23	1	f_6

četnosti statistického znaku f_i

Statistické zpracování dat: Zpracování statistického souboru

I. způsob Pokud jsou hodnoty souboru dat uvedeny jednotlivě, lze použít funkce **EXCEL**u

Momentové charakteristiky statisticky zpracovávaného souboru dat

- střední hodnota $\bar{x}$ = **Průměr(seznam_hodnot)**
- rozptyl σ^2 = **Var.P(seznam_hodnot)**, směrodatná odchylka
  σ = **SmOdch.P(seznam_hodnot)**
  nebo σ = **Var.P(seznam_hodnot)^(1/2)**
- koeficient šikmosti A = **Skew.P(seznam_hodnot)**
- koeficient špičatosti $e \approx$ **Kurt(seznam_hodnot)**

Kvantilové charakteristiky statisticky zpracovávaného souboru dat

- medián =2. kvartil
- kvartily $x_{0,25}$ = **Quartil.inc(seznam_hodnot ; 1)** $x_{0,5}$ = **Quartil.inc(seznam_hodnot ; 2)** $x_{0,75}$ = **Quartil.inc(seznam_hodnot ; 3)**
- decily =každý desátý percentil
- percentily. $x_{0,01}$ = **Percentil.inc(seznam_hodnot ; 0,01)** $x_{0,21}$ = **Percentil.inc(seznam_hodnot ; 0,21)** $x_{0,06}$ = **Percentil.inc(seznam_hodnot ; 0,06)**

Typickou hodnotou souboru zpracovávaných dat je ta nejčastější

- modus. Mo = **Mode.sngl(seznam_hodnot)**

Statistické zpracování dat: Zpracování statistického souboru

II. způsob Pokud jsou hodnoty souboru dat uvedeny s četnostmi, používáme vzorce

Momentové charakteristiky statisticky zpracovávaného souboru dat

- **střední hodnota** $\bar{x} = \frac{1}{N}\sum_i x_i \cdot f_i$
- **rozptyl** $\sigma^2 = \frac{1}{N}\sum_i (x_i - \bar{x})^2 \cdot f_i$ **směrodatná odchylka** $\sigma = \sqrt{\sigma^2}$
- **koeficient šikmosti** $A = \frac{1}{\sigma^3} \cdot \frac{1}{N}\sum_i (x_i - \bar{x})^3 \cdot f_i$
- **koeficient špičatosti** $e = \frac{1}{\sigma^4} \cdot \frac{1}{N}\sum_i (x_i - \bar{x})^4 \cdot f_i - 3$

Kvantilové charakteristiky statisticky zpracovávaného souboru dat SEŘAZANÉHO vzestupně

- **kvartily**, právě 25%, 50%, 75% pozorovaných měření má hodnotu znaku pod hranicí $x_{0,25}$, $x_{0,5}$, $x_{0,75}$
- **decily**, právě 10%, 20%, 30 % …. 90% pozorovaných měření má hodnotu znaku pod hranicí $x_{0,1}$, $x_{0,2}$, $x_{0,3}$ ….. $x_{0,9}$
- **percentily** analogicky $x_{0,01}$, $x_{0,02}$, $x_{0,03}$ …..

Typickou hodnotou souboru zpracovávaných dat je ta nejčastější

- **modus** hodnota x_i statistického znaku **s největší četností** f_i