Náhodná veličina a rozdělení pravděpodobnosti

Diskrétní a spojitá náhodná veličina

SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

doc. Ing. Petr Seďa, Ph.D.

SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

Co se dnes dozvíte?

- Náhodná veličina, zákon rozdělení pravděpodobnosti.
- Diskrétní náhodná veličina.
- Pravděpodobnostní funkce.
- Spojitá náhodná veličina.
- Funkce hustoty pravděpodobnosti a distribuční funkce.

Co už možná znáte?

Paretův záklon (např. rozdělení majetku):

Cca 20% obyvatel vlastní 80% bohatství, naopak 80% občanů se dělí pouze o 20% majetku.

Nelze určit, jaký majetek má náhodně vybraný občan, ale rozdělení majetku mezi všemi občany je možné vyjádřit matematickou funkcí.

⬇

Bohatství občana je náhodnou veličinou.

Náhodná veličina

Intuitivní definice:

- číselná veličina, jejíž hodnota je jednoznačně určena výsledkem náhodného pokusu

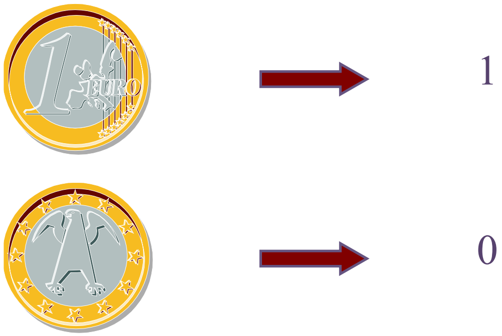
X – životnost zakoupeného výrobku
Y – počet zákazníků v obchodě
Z – kurs eura vůči koruně

Náhodná veličina

Definice náhodné veličiny:

- každému výsledku pokusu můžeme přiřadit reálné číslo, které jej bude charakterizovat

Náhodnou veličinou *X* nazveme zobrazení, které každému možnému výsledku pokusu přiřadí konkrétní reálné číslo.



Náhodná veličina

Realizace náhodné veličiny:

***X* – náhodná veličina**

obor hodnot náhodné veličiny *X*

množina všech reálných čísel, kterých může náhodná veličina X nabýt

realizace náhodné veličiny *X*

hodnota přiřazená výsledku konkrétního pokusu

$X = 0$

Diskrétní a spojitá náhodná veličina

Typy náhodných veličin:

- **diskrétní náhodná veličina**

  - nabývá spočetně mnoha hodnot (konečně nebo nekonečně)

  Příklady: první číslo vytažené z osudí ve Sportce, životnost konkrétního výrobku ve dnech

- **spojitá náhodná veličina**

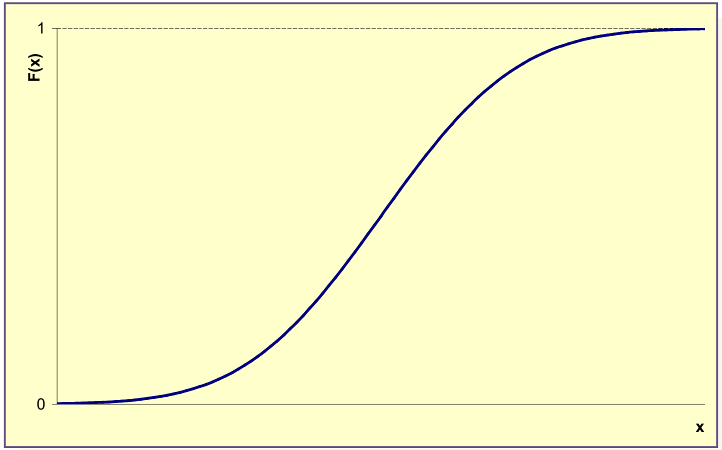
  - nabývá libovolnou hodnotu z daného intervalu

  Příklady: kurs eura vůči koruně dle ČNB v daném dni, ekonomický výsledek firmy v daném roce

Distribuční funkce spojité náhodné veličiny:

distribuční funkce $F(x)$ – pravděpodobnost na intervalu

$$F(x) = P(X \leq x)$$



neklesající
začíná v 0
končí v 1

$$a < b \;\Rightarrow\; F(a) \leq F(b)$$

$$\lim_{x \to -\infty} F(x) = 0$$

$$\lim_{x \to +\infty} F(x) = 1$$

SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

Distribuční funkce diskrétní náhodné veličiny:

- někdy se zavádí distribuční funkce i pro diskrétní náhodnou veličinu

1. varianta: $F(x) = P(X \leq x)$

2. varianta: $F(x) = P(X < x)$

- obvykle si vystačíme pouze s pravděpodobnostní funkcí

Diskrétní náhodná veličina

Rozdělení diskrétní náhodné veličiny:

- zákon rozdělení pravděpodobnosti náhodné veličiny

- rozdělení diskrétní náhodné veličiny

pravděpodobnostní funkce $p(x)$

- přiřazuje hodnotě diskrétní náhodné veličiny pravděpodobnost výskytu

$$\boxed{p(x) = P(X = x)}$$

tabulka

graf

funkční předpis

Příklad 1 – výstupní kontrola

Výrobek má 10% zmetkovitost. V soupravě jsou vždy 3 výrobky z dané série. Vybereme náhodně jednu soupravu a otestujeme všechny výrobky. Náhodná veličina X bude počet vadných výrobků v soupravě.

a) Určete **obor hodnot proměnné *X*,**
b) A také **pravděpodobnostní funkci.**

obor hodnot proměnné $X = \{0, 1, 2, 3\}$

$$\left.\begin{aligned} p(0) &= P(X = 0) = 0{,}9^3 = 0{,}729 \\ p(1) &= P(X = 1) = 3 \cdot 0{,}9^2 \cdot 0{,}1 = 0{,}243 \\ p(2) &= P(X = 2) = 3 \cdot 0{,}9 \cdot 0{,}1^2 = 0{,}027 \\ p(3) &= P(X = 3) = 0{,}1^3 = 0{,}001 \end{aligned}\right\} \sum_i p(x_i) = 1$$

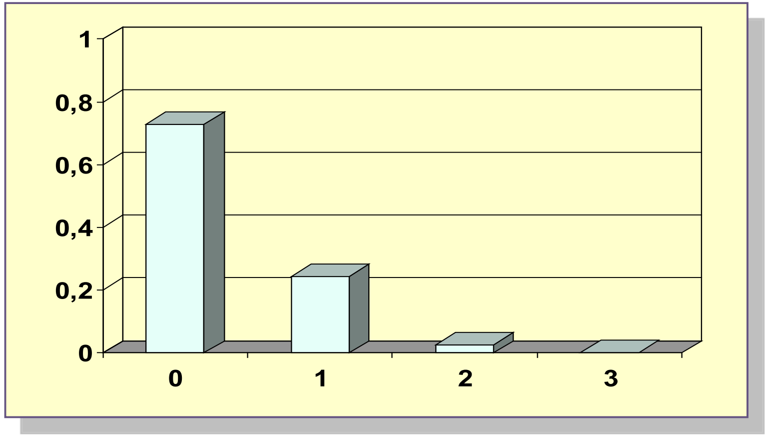
Příklad 1 – výstupní kontrola

SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

TABULKA:

x	0	1	2	3
$p(x)$	0,729	0,243	0,027	0,001

GRAF:



VZOREC:
počkejte si na příští přednášku

Příklad 2 – kvalita výrobku

Při ověřování kvality výroby v podniku Opavia jsou náhodně vybrány dva výrobky a je testována jejich kvalita. Počet vadných výrobků mezi vybranými modelujeme náhodnou veličinou X. Z dlouhodobého pozorování jsou známy údaje uvedené v následující tabulce.

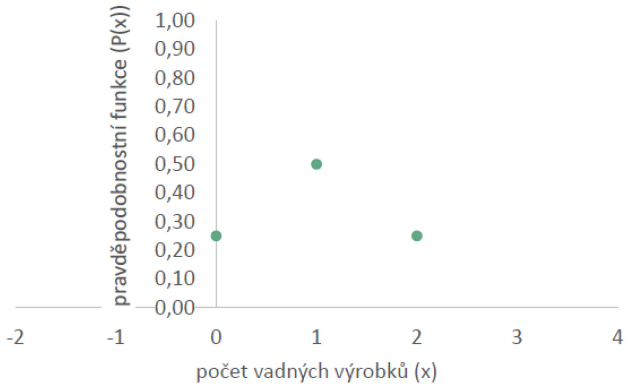
Počet vadných výrobků	Pravděpodobnost
0	0,25
1	0,50
2	0,25

a) určete **pravděpodobnostní funkci** počtu vadných výrobků v testovaném vzorku,
b) určete **distribuční funkci** počtu vadných výrobků v testovaném vzorku.

Příklad 2 – kvalita výrobku

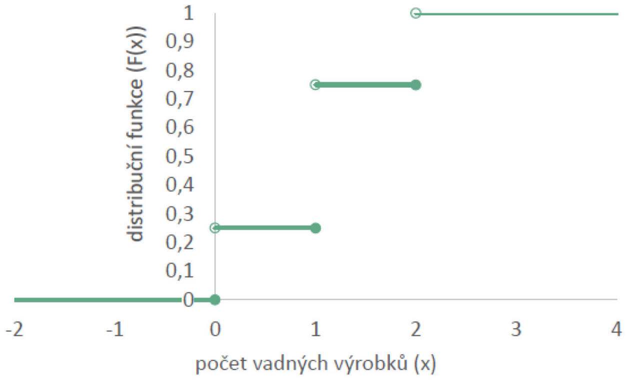
Ad a)

Počet vadných výrobků x	Pravděpodobnost $P(x)$
0	0,25
1	0,50
2	0,25
Celkem	**1,00**



Ad b)

Počet vadných výrobků x	Distribuční funkce $F(x) = P(X < x)$
$(-\infty; 0\rangle$	0,00
$(0; 1\rangle$	0,25
$(1; 2\rangle$	0,75
$(2;\ \infty)$	1,00



Rozdělení spojité náhodné veličiny:

u spojité veličiny nemá smysl určovat bodovou pravděpodobnost typu $P(X = x)$

Jaká je pravděpodobnost, že zítřejší kurs eura u ČNB bude přesně 32,50 Kč / € ?

NEMÁ SMYSL ($P \to 0$)

Jaká je pravděpodobnost, že zítřejší kurs eura u ČNB bude mezi 32 a 33 Kč / € ?

LZE URČIT VÝPOČTEM

SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

Pravděpodobnost na intervalu:

pro spojitou náhodnou veličinu ***X*** platí:

$$\boxed{F(a < X < b) = F(b) - F(a)}$$

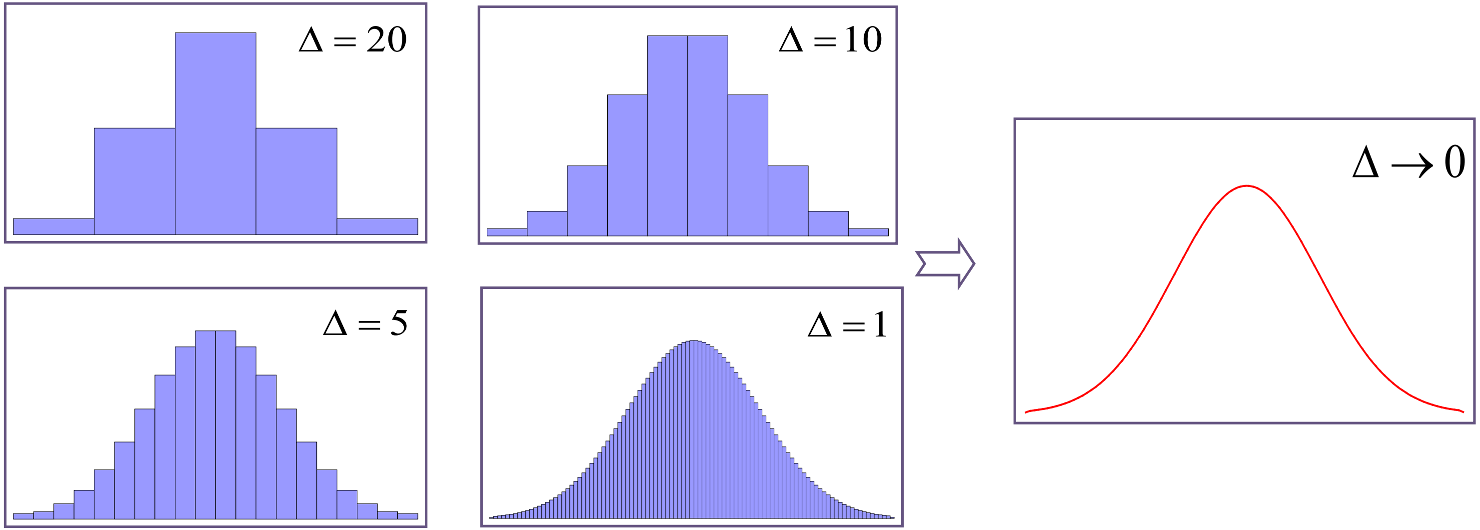
nezáleží na typu intervalu:

$$\left.\begin{array}{l} F(a < X \leq b) \\ F(a \leq X < b) \\ F(a \leq X \leq b) \end{array}\right\} = F(b) - F(a)$$

Spojitá náhodná veličina

Pojem hustota pravděpodobnosti:

Intervalové rozdělení výšky dospělých mužů   Δ – šířka intervalu v cm



Čím užší intervaly děláme, tím více se hodnoty grafu blíží spojité křivce.
Tato křivka vyjadřuje **hustotu pravděpodobnosti** daného rozdělení.

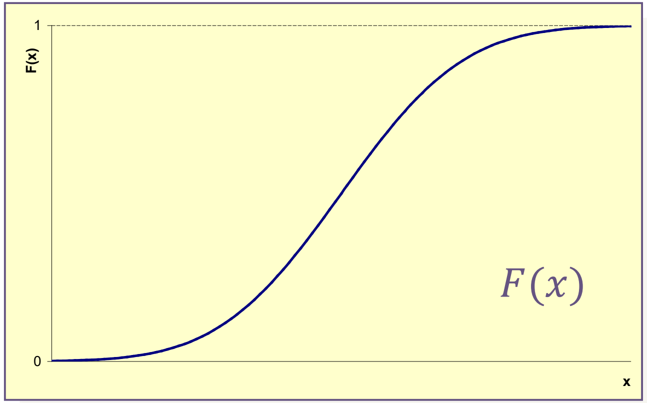
Spojitá náhodná veličina

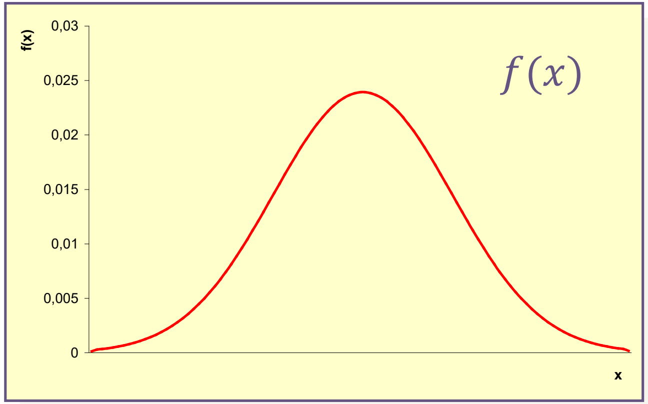
Funkce hustoty pravděpodobnosti $f(x)$:

frekvenční funkce spojité náhodné veličiny

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

$$f(x) = \frac{dF(x)}{dx}$$





Spojitá náhodná veličina

Vlastnosti spojité náhodné veličiny:

Distribuční funkce $F(x)$ je omezená zdola (0) i shora (1).
Frekvenční funkce $f(x)$ je omezená pouze zdola (0).

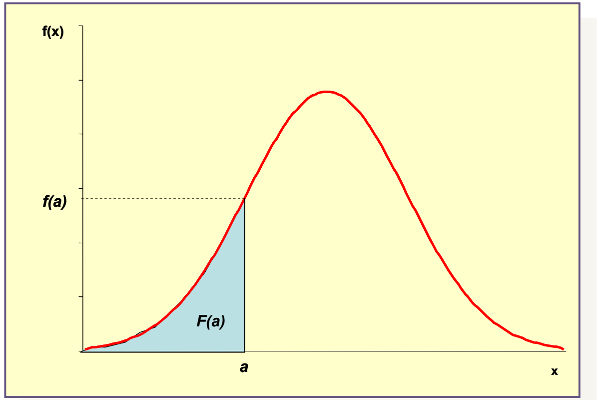
Distribuční funkce $F(x)$ „začíná“ v 0 a „končí“ v 1.
Frekvenční funkce $f(x)$ „začíná“ i „končí“ v 0.

Distribuční funkce $F(x)$ je neklesající.
Frekvenční funkce $f(x)$ může růst i klesat.

Obsah plochy pod grafem frekvenční funkce $f(x)$ **je roven 1.**

$$F(a < X < b) = F(b) - F(a)$$

$$F(a < X < b) = \int_a^b f(x)dx$$



Příklad 3 – trolejbusy

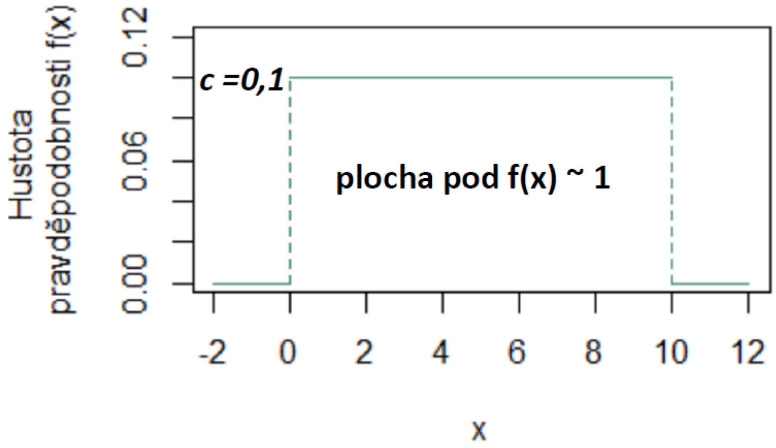
Trolejbusy v Opavě jezdí v pravidelných intervalech po 10 minutách. Cestující přijde na zastávku v libovolném okamžiku. Náhodná veličina X představuje dobu čekání na příjezd trolejbusu. Rozdělení dané náhodné veličiny je dáno **hustotou pravděpodobností**, která je v intervalu 0 - 10 minut konstantní, mimo tento interval je nulová, tj.: $f(x) = \begin{cases} c, \text{pro } x \in \langle 0; 10 \rangle \\ 0, \text{pro } x \neq \langle 0; 10 \rangle \end{cases}$.

a) Určete konstantu c.
b) Určete distribuční funkci $F(x)$.
c) Určete pravděpodobnost, že cestující bude čekat nejvýše 5 minut, alespoň 3 minuty, právě 7 minut.

Příklad 3 – trolejbusy

Trolejbusy v Opavě jezdí v pravidelných intervalech po 10 minutách. Cestující přijde na zastávku v libovolném okamžiku. Náhodná veličina X představuje dobu čekání na příjezd trolejbusu. Rozdělení dané náhodné veličiny je dáno **hustotou pravděpodobností**, která je v intervalu 0 - 10 minut konstantní, mimo tento interval je nulová, tj.: $f(x) = \begin{cases} c, \text{pro } x \in \langle 0; 10 \rangle \\ 0, \text{pro } x \neq \langle 0; 10 \rangle \end{cases}$.

a) Určete konstantu c.



Příklad 3 – trolejbusy

Trolejbusy v Opavě jezdí v pravidelných intervalech po 10 minutách. Cestující přijde na zastávku v libovolném okamžiku. Náhodná veličina X představuje dobu čekání na příjezd trolejbusu. Rozdělení dané náhodné veličiny je dáno **hustotou pravděpodobností**, která je v intervalu 0 - 10 minut konstantní, mimo tento interval je nulová, tj.: $f(x) = \begin{cases} c, \text{pro } x \in \langle 0; 10 \rangle \\ 0, \text{pro } x \neq \langle 0; 10 \rangle \end{cases}$.

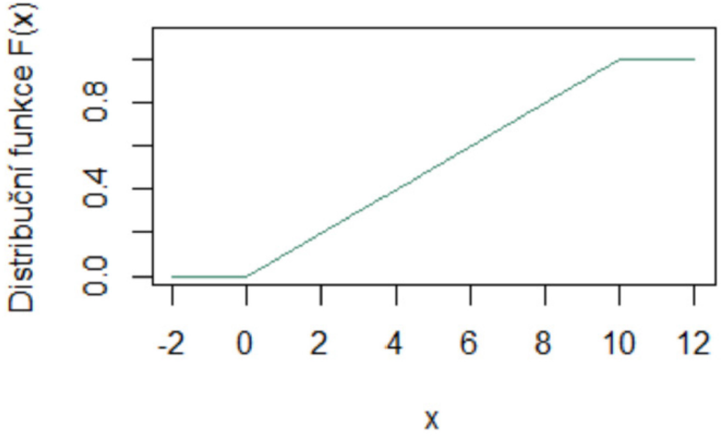
SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

b) Určete distribuční funkci $F(x)$.

$$F(x) = \int_{-\infty}^{x} f(t)\, \mathrm{d}t$$

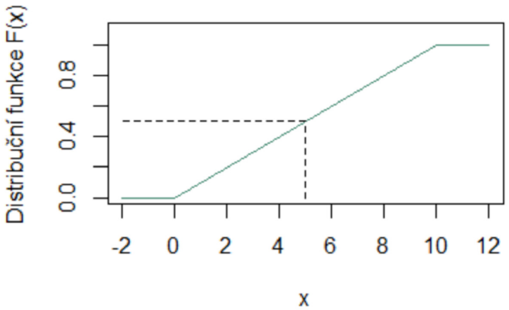
$$F(x) = \begin{cases} 0 & pro\ x < 0 \\ \int_{-\infty}^{0} 0\, \mathrm{d}t + \int_{0}^{x} 0{,}1\, \mathrm{d}t & pro\ 0 \leq x \leq 10 \\ 1 & pro\ x > 10 \end{cases}$$

$$F(x) = \begin{cases} 0 & pro\ x < 0 \\ 0{,}1x & pro\ 0 \leq x \leq 10 \\ 1 & pro\ x > 10 \end{cases}$$



Příklad 3 – trolejbusy

Trolejbusy v Opavě jezdí v pravidelných intervalech po 10 minutách. Cestující přijde na zastávku v libovolném okamžiku. Náhodná veličina X představuje dobu čekání na příjezd trolejbusu. Rozdělení dané náhodné veličiny je dáno **hustotou pravděpodobností**, která je v intervalu 0 - 10 minut konstantní, mimo tento interval je nulová, tj.: $f(x) = \begin{cases} c, \text{pro } x \in \langle 0; 10 \rangle \\ 0, \text{pro } x \neq \langle 0; 10 \rangle \end{cases}.$



c) Určete pravděpodobnost, že cestující bude čekat:

i. nejvýše 5 minut,

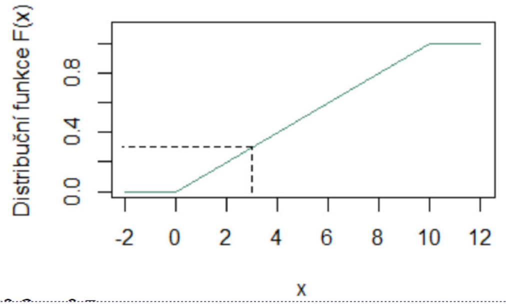
ii. alespoň 3 minuty,

iii. právě 7 minut.

Ad i. $P(X \leq 5) = P(X < 5) = \int_{-\infty}^{5} f(x)\mathrm{d}x = F(5) = 0{,}5.$

Ad ii. $P(X \geq 3) = 1 - P(X < 3) = 1 - F(3) = 1 - 0{,}3 = 0{,}7.$

Ad iii. $P(X = 7) = 0$ (Pravděpodobnostní funkce SNV je ve všech bodech nulová.)



SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

Vybrané charakteristiky náhodné veličiny:

- charakteristiky polohy
  - střední hodnota $E(X)$
  - medián $Me(X)$
  - modus $Mo(X)$

- charakteristiky variability
  - rozptyl $D(X)$
  - směrodatná odchylka $\sigma(X)$

- kvantily

Charakteristiky náhodné veličiny

Střední hodnota náhodné veličiny $E(X)$:

- $E(X)$ – expected value (očekávaná hodnota)

- pro diskrétní náhodnou veličinu

$$\boxed{E(X) = \sum_i x_i \cdot p(x_i)}$$

- pro spojitou náhodnou veličinu

$$\boxed{E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx}$$

meze nahradíme skutečnými hranicemi náhodné veličiny

SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

Medián a modus náhodné veličiny:

- pro diskrétní náhodnou veličinu:

$\boldsymbol{Me(X)}$ – nejednoznačné, obvykle se neurčuje
$\boldsymbol{Mo(X)}$ – hodnota s největší pravděpodobnostní funkcí

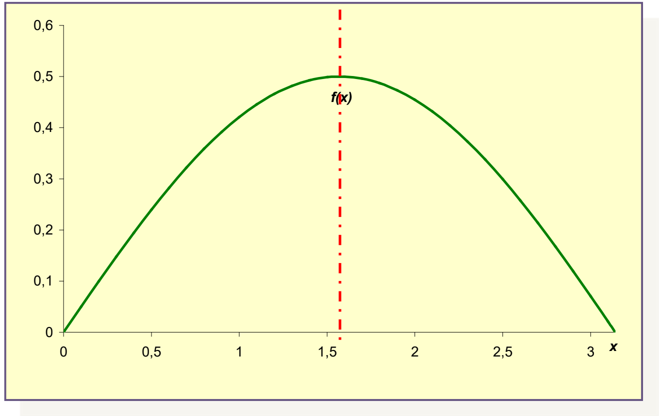
- pro spojitou náhodnou veličinu:

$\boldsymbol{Me(X)}$ – hodnota s distribuční funkcí rovnou 0,5
$\boldsymbol{Mo(X)}$ – hodnota s největší hustotou pravděpodobnosti

SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

Medián, modus a střední náhodné veličiny:

Co lze říci o tvaru rozdělení náhodné veličiny X, která má tyto shodné charakteristiky: střední hodnotu $E(X)$, medián $Me(X)$ i modus $Mo(X)$?



Frekvenční funkce veličiny X je symetrická kolem své střední hodnoty $E(X)$.

Charakteristiky náhodné veličiny

Rozptyl náhodné veličiny $D(X)$:

$D(X)$ – dispersion (rozptyl)

pro diskrétní náhodnou veličinu

$$\boxed{D(X)=\sum_i x_i^2\cdot p(x_i)-E^2(X)}$$

pro spojitou náhodnou veličinu

$$\boxed{D(X)=\int_{-\infty}^{+\infty} x^2\cdot f(x)dx-E^2(X)}$$

Charakteristiky náhodné veličiny

Směrodatná odchylka $\sigma(X)$:

pro diskrétní i spojitou náhodnou veličinu

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

SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

Kvantily spojité náhodné veličiny:

kvantil – inverzní funkce k distribuční funkci, dělí pravděpodobnostní prostor na dvě části s pravděpodobnostmi $\boldsymbol{p}$ a $\mathbf{1}-\boldsymbol{p}$

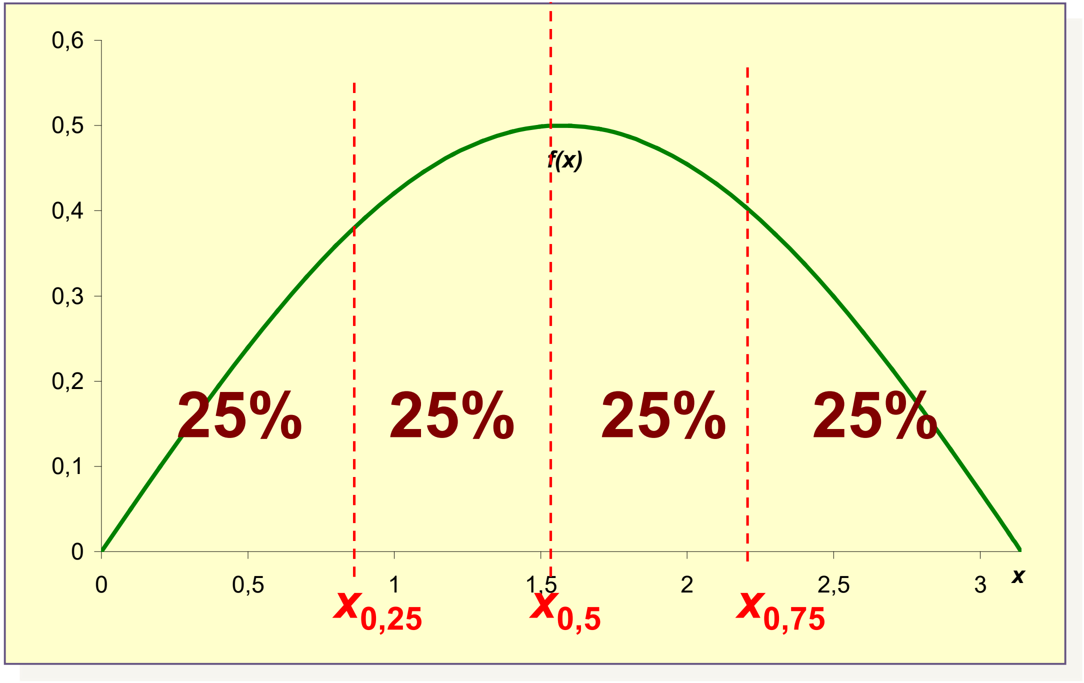
$$\boxed{x_p \approx F^{-1}(x)}$$

p–kvantil x_p získáme řešením rovnice:

$$\boxed{F(x_p) = p}$$

SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

Kvantily spojité náhodné veličiny:



Příklad 4 – kvalita výrobku

Při ověřování kvality výroby v podniku Opavia jsou náhodně vybrány dva výrobky a je testována jejich kvalita. Počet vadných výrobků mezi vybranými modelujeme náhodnou veličinou X. Z dlouhodobého pozorování jsou známy údaje uvedené v následující tabulce.

Počet vadných výrobků	Pravděpodobnost
0	0,25
1	0,50
2	0,25

Určete:

a) střední hodnotu počtu vadných výrobků v testovaném vzorku,

b) rozptyl počtu vadných výrobků v testovaném vzorku,

c) směrodatnou odchylku počtu vadných výrobků v testovaném vzorku,

Příklad 4 – kvalita výrobku

SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

Ad a) střední hodnota počtu vadných výrobků v testovaném vzorku:

x	$P(x)$	$x \cdot P(x)$
0	0,25	0,00
1	0,50	0,50
2	0,25	0,50
Celkem	**1,00**	**1,00**

$$E(X) = \mu = \sum_{(i)} x_i \cdot P(x_i) = 1.$$

Ad b) rozptyl počtu vadných výrobků v testovaném vzorku:

x	$P(x)$	$x \cdot P(x)$	$x^2 \cdot P(x)$
0	0,25	0,00	0,00
1	0,50	0,50	0,50
2	0,25	0,50	1,00
Celkem	**1,00**	**1,00**	**1,50**

$$D(X) = \sum_i x_i^2 \cdot p(x_i) - \left(E(X)\right)^2 = 1{,}50 - 1^2 = 0{,}50.$$

Rozptyl počtu vadných výrobků v testovaném vzorku je 0,5 ks^2.

Příklad 4 – kvalita výrobku

Ad c) směrodatná odchylka počtu vadných výrobků v testovaném vzorku:

$$\sigma(X) = \sqrt{D(X)} = \sqrt{0{,}50} \cong 0{,}71.$$

Směrodatná odchylka počtu vadných výrobků v testovaném vzorku je 0,71 ks.

Literatura

1. Janáček J. *Statistika jednoduše*. Praha: Grada, 2022. **(kapitola 2).**

2. Ramík J. a Čemerková Š. *Statistika A*. Karviná: SLU, 2000. **(kapitola 3).**

SLEZSKÁ UNIVERZITA
FAKULTA VEŘEJNÝCH POLITIK V OPAVĚ

Děkuji za pozornost.