29 Pravděpodobnost a statistika – met.

Stručný přehled teorie

Pravděpodobnost

Náhodný pokus – pokus, jehož výsledek záleží i při dodržení předem stanovených podmínek na náhodě.

$\Omega=\{\omega_1; \omega_2;\dots; \omega_n\}$ … množina všech možných výsledků náhodného pokusu

Náhodný jev A je podmnožinou množiny Ω ..... $A \subseteq \Omega$

Klasická (Laplaceova) definice pravděpodobnosti:

Nechť náhodný pokus splňuje předpoklady:

1) Všech možných výsledků je konečný počet
2) Všechny výsledky mají stejnou šanci na realizaci
3) Všechny výsledky se navzájem vylučují (tj. žádné dva nemohou nastat současně)
4) Jeden z výsledků jistě nastane.

Pak pravděpodobností jevu A se nazývá číslo $\boxed{P(A) = \frac{m}{n}}$, kde n je počet všech možných výsledků (počet prvků Ω) a m je počet výsledků příznivých jevu A (počet prvků A).

Některé vlastnosti pravděpodobnosti:

1. a) *Jistý jev* … $A = \Omega$ => $P(A)=\frac{n}{n}=1$

b) *Nemožný jev* … $A = \emptyset$ => $P(A)=\frac{0}{n}=0$

c) *Náhodný jev* … $A \subseteq \Omega$ => $0 \le P(A) \le 1$

2. Opačný jev k jevu A je takový jev A‘, který nastává právě tehdy, když nenastal jev A.

Tedy: $A' = \Omega - A$ (přesněji $A \cap A' = \emptyset \wedge A \cup A' = \Omega$)

Pak $\boxed{P(A') = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P(A)}$

3. Sjednocení jevů A, B je jev $A \cup B$, který nastane právě tehdy, když nastane aspoň jeden z jevů A nebo B.

a) Platí-li, že se jevy A, B navzájem vylučují (tj. $A \cap B = \emptyset$), pak $\boxed{P(A\cup B) = P(A) + P(B)}$

b) Pokud se jevy A, B navzájem nevylučují (tj. $A \cap B \neq \emptyset$), pak $\boxed{P(A\cup B) = P(A) + P(B) - P(A\cap B)}$

4. Průnik jevů A, B je jev $A \cap B$, který nastane právě tehdy, když nastane jev A a zároveň jev B. Jestliže A, B jsou **nezávislé jevy**, pak $\boxed{P(A\cap B) = P(A).P(B)}$. Jsou-li A, B závislé jevy, pak $P(A\cap B) \neq P(A).P(B)$

Statistická (zobecněná) definice pravděpodobnosti:

Pravděpodobnost P(A) jevu A je určena přibližně jeho relativní četností při dostatečně velkém počtu opakování náhodného pokusu.

Nechť $\Omega = \{\omega_1; \omega_2; \ldots; \omega_n\}$ je množina všech možných výsledků náhodného pokusu

a $p_1, p_2, \ldots, p_n$ jsou jejich relativní četnosti (tzn. $\sum_{i=1}^{n} p_i = 1$).

Pak $\boxed{P(A) = \sum_{i=1}^{k} p_{j_i}}$, kde p_{j_i} značí relativní četnost výsledku $\omega_{j_i} \in A$ (*k* je počet prvků A).

Podmíněná pravděpodobnost – pravděpodobnost jevu A podmíněnou jevem B určíme takto:

$$\mathbf{P(A|B)} = \frac{P(A \cap B)}{P(B)}$$

Bernoulliho schéma:

Nechť při n-násobném opakování náhodného pokusu je stále stejná pravděpodobnost zdaru *p*

a pravděpodobnost nezdaru *q* (tedy $q = 1 - p$). Pak pravděpodobnost jevu A_k, že zdar nastane

v těchto n pokusech právě k-krát je dána vztahem: $\boxed{P(A_k) = \binom{n}{k} . p^k . q^{n-k}}$

Základy statistiky

Statistika se zabývá zkoumáním a zpracováním velkého množství dat souvisejících s hromadnými jevy.

Základní pojmy

- **statistický soubor** – konečná neprázdná množina objektů, které zkoumáme (např. obyvatelé Brna, obyvatelé ČR, rodinní příslušníci zaměstnanců určité firmy, dopravní nehody v určité oblasti za určité období, výrobky vyrobené v určité firmě za určité období ,…)
- **statistická jednotka** – prvek statistického souboru (např. jeden určitý obyvatel, jeden daný výrobek)
- **rozsah statistického souboru** – počet prvků statistického souboru
- **statistický znak** – společná vlastnost prvků statistického souboru, kterou zjišťujeme (např. věk, národnost, výše měsíčního příjmu, výška postavy, kvalita výrobku (vadný nebo bez vady), …);

znak může být

- ***kvantitativní*** (číselný) – např. počet obyvatel daného věku, výše škody při nehodě, …

- ***kvalitativní*** (popsán slovy) – např. povolání, druh nemoci, příčina dopravní nehody, …

Pozn.: 1) *kvalitativní* znak může mít někdy více možností (např. příčin nehody může být víc) – pak se musí vybrat jedna, která je hlavní (ostatní mohou tvořit kategorii „jiné“);

Pozn.: 2) *nejjednodušší kvalitativní* znak je znak *alternativní* – dán jevem a jeho opakem – např. voják-nevoják, muž-žena, plavec-neplavec, prospěl-neprospěl, …

- **absolutní četnost** hodnoty znaku x_i – číslo n_i udávající počet prvků daného statistického souboru, které vykazují sledovanou hodnotu x_i, neboli udávající, pro kolik prvků souboru nabývá statistický znak určité hodnoty nebo rozmezí hodnot (např. kolik nezaměstnaných osob je evidováno v dané oblasti, kolik osob má měsíční příjem ve vybraném rozmezí, …)

- **relativní četnost** znaku – poměr $\frac{n_i}{n}$ absolutní četnosti dané hodnoty a rozsahu souboru

  Pozn.: Relativní četnost se nejčastěji uvádí v procentech $\frac{n_i}{n}.100\ \%$

Statistické soubory rozdělujeme na
- **základní** (mohou mít pro zkoumání příliš velký rozsah)
- **výběrové** (část základního souboru, na němž se provádí zkoumání)

Charakteristiky statistického souboru

1. **Charakteristiky polohy** hodnot znaku jsou číselné hodnoty, které určitým způsobem charakterizují typickou hodnotu sledovaného znaku

   - **Aritmetický průměr** – součet všech hodnot zjištěných znaků dělených jejich počtem.
     $\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$ nebo tzv. vážený průměr $\boxed{\bar{x} = \frac{1}{n}\sum_{j=1}^{r} x_j n_j}$
   - **Modus** **Mod(x) – hodnota** znaku s největší četností.
   - **Medián** **Med(x) – je** prostřední hodnota znaku, jsou-li hodnoty uspořádány podle velikosti.
     $\boxed{Med(x) = x_{\frac{n+1}{2}}}$, je-li n liché, $\boxed{Med(x) = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)}$, je-li n sudé.
   - **Geometrický průměr** $\boxed{\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots\ldots \cdot x_n}}$

2. **Charakteristiky variability**
   Každá charakteristika polohy je číslo, kolem něhož jednotlivé hodnoty znaku kolísají. Charakteristiky variability vyjadřují „velikost“ onoho kolísání.

   - **Rozptyl** s_x^2 se definuje jako průměr druhých mocnin odchylek od aritmetického průměru
     $$\boxed{s_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{1}{n}\sum_{j=1}^{r}\left(x_j^* - \bar{x}\right)^2 n_j}$$
   - **Směrodatná odchylka** $\boxed{s_x = \sqrt{s_x^2}}$
   - **Variační koeficient** v_x – podíl směrodatné odchylky a aritmetického průměru – udává se v procentech $\boxed{v_x = \frac{s_x}{\bar{x}}.100\,\%}$

- **Koeficient korelace** r - souvisí s tím, že se velmi často zkoumá, zda a jak jsou na sobě závislé dva znaky *x* a *y*. Koeficient korelace vyjadřuje míru vzájemné závislosti těchto znaků *x* a *y*.

$$\boxed{r = \frac{k}{s_x . s_y}}, \qquad \text{kde } k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) ,$$

$$s_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \;, \quad s_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Pozn.: Vždy platí: $|r| \le 1$. Čím víc se hodnota *r* blíží k **1**, tím považujeme závislost *x* a *y* za silnější.

Met.: Když učitel probírá teorii k tématu Pravděpodobnost, přímo se mu nabízejí velmi jednoduché konkrétní situace, kterými by měl průběžně doplňovat svůj výklad, aby i nejslabší studenti okamžitě pochopili probírané pojmy, děje, jevy. Určitě k nejvhodnějším takovým situacím patří házení kostky ze hry Člověče, nezlob se.

Náhodný pokus	Hod kostkou
Množina Ω	$\Omega = \{1, 2, 3, 4, 5, 6\}$
Náhodný jev A	$A \subseteq \Omega$, např. $A = \{1, 2\}$ ... padne číslo menší než 3 - splněny předpoklady pro použití Laplaceovy definice pravděpodobnosti: 1) konečný počet prvků množiny Ω, 2) stejná šance pro všechny, 3) dva výsledky nemohou nastat současně, 4) jeden výsledek nastane jistě. $P_{(A)} = \frac{m}{n} = \frac{\lvert A\rvert}{\lvert\Omega\rvert} = \frac{2}{6} = \frac{1}{3}$
Jistý jev B	B ...padne číslo menší než 8 $\quad B = \Omega = \{1, 2, 3, 4, 5, 6\}$ $\boxed{P_{(B)}} = \frac{m_B}{n} = \frac{\lvert B\rvert}{\lvert\Omega\rvert} = \frac{6}{6} = \boxed{1}$
Nemožný jev C	C ... padne číslo větší než 6 $\quad C = \emptyset$ $\boxed{P_{(C)}} = \frac{m_C}{n} = \frac{\lvert C\rvert}{\lvert\Omega\rvert} = \frac{0}{6} = \boxed{0}$
Náhodný jev D	D ... padne sudé číslo $\quad D = \{2, 4, 6\}$ $\boxed{P_{(D)}} = \frac{m_D}{n} = \frac{\lvert D\rvert}{\lvert\Omega\rvert} = \frac{3}{6} = \frac{1}{2}$ $\quad \boxed{0 \le P_{(D)} \le 1}$
Opačný jev A´ k jevu A	A ... padne číslo menší než 3 $\quad A = \{1, 2\}$ A´... padne číslo větší nebo rovné 3 $\quad A = \{3, 4, 5, 6\}$

Platí: $(A \cap A' = \emptyset) \;\wedge\; (A \cup A' = \Omega)$.

$P_{(A)} = \frac{2}{6} = \frac{1}{3}; \quad P_{(A')} = \frac{4}{6} = \frac{2}{3}.$ $\qquad \boxed{P_{(A)} + P_{(A')} = 1}$

Sjednocení jevů

a)

A ... padne číslo menší než 3 $A = \{1, 2\}$

E ... padne číslo 5 $E = \{5\}$

$A \cup E = \{1, 2, 5\}$

$P_{(A)} = \frac{2}{6}; \quad P_{(E)} = \frac{1}{6};$ $= \frac{m_{A \cup E}}{n} = \frac{3}{6} = \frac{2}{6} + \frac{1}{6} =$

b)

A ... padne číslo menší než 3 $A = \{1, 2\}$

D ... padne sudé číslo $D = \{2, 4, 6\}$

$A \cup D = \{1, 2, 4, 6\} \quad A \cap D = \{2\}$

$P_{(A)} = \frac{2}{6}; \quad P_{(D)} = \frac{3}{6};$ $= \frac{m_{A \cup D}}{n} = \frac{4}{6} = \frac{2}{6} + \frac{3}{6} - \frac{1}{6} =$

A ... při prvním hodu padne 2 $P_{(A)} = \frac{1}{6}$

B ... při druhém hodu padne sudé číslo $P_{(B)} = \frac{1}{2}$

$A \cap B = \{[2,2], [2,4], [2,6]\} \quad m_{A \cap B} = 3$

$n = V'_{(2,6)} = 6^2 = 36$

$P_{(A \cap B)} = \frac{m_{A \cap B}}{n} = \frac{3}{36} = \frac{1}{12} \qquad P_{(A)} . P_{(B)} = \frac{1}{6} . \frac{1}{2} = \frac{1}{12}$

Jevy A, B jsou nezávislé.

A ... při prvním hodu padne 2 $P_{(A)} = \frac{1}{6}$

C ... součet při dvou hodech bude 5

5: 1+4, 4+1, 2+3, 3+2

$P_{(C)} = \frac{4}{36} = \frac{1}{9}$

$A \cap C = \{[2,3]\} \quad m_{A \cap C} = 1$

$n = V'_{(2,6)} = 6^2 = 36$

$P_{(A \cap B)} = \frac{m_{A \cap C}}{n} = \frac{1}{36} \qquad P_{(A)} . P_{(B)} = \frac{1}{6} . \frac{1}{9} = \frac{1}{54}$

Jevy A, B jsou závislé.

Statistická (zobecněná) definice pravděpodobnosti

- využívá se v případě, že všechny výsledky (prvky množiny Ω) nemají stejnou šanci, aby nastaly.

Př.: Házení nehomogenní kostkou na Člověče, nezlob se. S jakou pravděpodobností padne číslo větší než 3?

ω_i	n_i	p_i
1	2	$\frac{2}{30} = \frac{1}{15}$
2	4	$\frac{4}{30} = \frac{2}{15}$
3	5	$\frac{5}{30} = \frac{1}{6}$
4	6	$\frac{6}{30} = \frac{1}{5}$
5	6	$\frac{6}{30} = \frac{1}{5}$
6	7	$\frac{7}{30}$
	$\Sigma n_i = 30$	$\Sigma p_i = 1$

$A = \{4, 5, 6\}$

$$P_{(A)} = \frac{1}{5} + \frac{1}{5} + \frac{7}{30} = \frac{19}{30}$$

Bernoulliho schéma

Př.: 20 hodů kostkou. Jev A … číslo 5 padne právě 3krát. Jaká je pravděpodobnost jevu A?

$$P_{(A)} = \binom{20}{3} . \left(\frac{1}{6}\right)^3 . \left(\frac{5}{6}\right)^{17} = \frac{20.19.18}{3.2.1} . \frac{5^{17}}{6^{20}} = 190 . \frac{5^{17}}{6^{19}}$$

Základy statistiky:

Př. 1: V souboru A byl sledován údaj o počtu dětí v 13 rodinách. Rozsah souboru je n = 13 (liché):

počet dětí	0	1	2	3	4	5	6	7
četnost	2	4	3	2	1	1	0	0

Aritmetický průměr$\overline{x} = \frac{1}{13}\sum_{i=1}^{13} x_i$=$\frac{0+0+1+1+1+1+2+2+2+3+3+4+5}{13} = \frac{25}{13} = 1{,}923$

(Vážený průměr $\overline{x} = \frac{1}{13}\sum_{j=1}^{8} x_j n_j$=$\frac{0.2+1.4+2.3+3.2+4.1+5.1+6.0+7.0}{13} = \frac{25}{13} = 1{,}923$)

Modus Mod(x) = 1 (má četnost 4, což je nejvíce)

Medián Med(x) = 2 (7. rodina).

Př. 2: V souboru B byl sledován údaj o počtu dětí ve 14 rodinách. Rozsah souboru je n = 14 (sudé):

počet dětí	0	1	2	3	4	5	6	7	8
četnost	2	5	3	1	1	0	0	1	1

Aritmetický průměr $\overline{x} = \frac{1}{14}\sum_{i=1}^{14} x_i$=$\frac{0+0+1+1+1+1+1+2+2+2+3+4+7+8}{14} = \frac{33}{14} = 2{,}357$

(Vážený průměr $\overline{x} = \frac{1}{14}\sum_{j=1}^{9} x_j n_j = \frac{0.2+1.5+2.3+3.1+4.1+5.0+6.0+7.1+8.1}{14} = \frac{33}{14} = 2{,}357$)

Modus Mod(x) = 1 (má četnost 5, což je nejvíce)

Medián Med(x) = 1,5 (aritmetický průměr ze 7. a 8.rodiny).

Základní poznatky:

1. Jaká je pravděpodobnost, že při hodu třemi kostkami padne součet 12?
   [0,116]
2. 55% populace tvoří ženy, 45% muži. Určitou chorobou trpí 1% žen a 5% mužů. Jaká je p., že náhodně vybraná osoba z populace trpí touto chorobou?
   [0,028]
3. Dvanáct studentů, mezi kterými je Pavel a Tomáš, mají ze svého středu vylosovat 4-člennou skupinu.
   Jaká je pravděpodobnost, že ve skupině bude:
   a) Tomáš b) Tomáš, ale Pavel ne c) Tomáš a Pavel d)Tomáš nebo Pavel
   [0,333 ; 0,242 ; 0,091 ; 0,576]
4. Jaká je pravděpodobnost, že se Jana a Tomáš narodili ve stejný měsíc?
   (Počítejte, že 1 měsíc je $^{1}/_{12}$ roku) [0,0833]

5. **Státní maturita 2017**

   Z 25 žáků jedné třídy domácí úkol 3 žáci nevypracovali, 6 žáků jej vypracovalo chybně a zbývající žáci jej vypracovali správně.

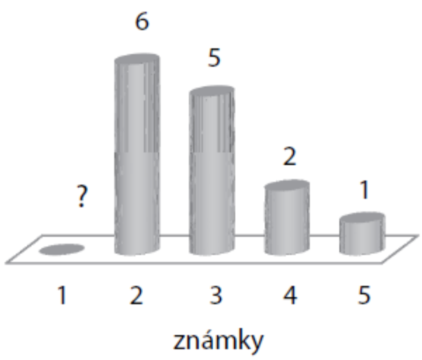
   Učitel náhodně vybere dvojici žáků.

   (CZVV)

   Jaká je pravděpodobnost, že oba vybraní žáci budou mít úkol vypracován správně? [2/5]

6. **Státní maturita 2015**

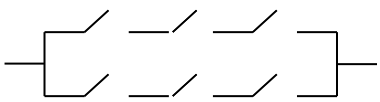
   Graf udává **četnost známek** z písemné práce, avšak počet jedniček není uveden. **Medián** je 2,5.



   Kolik písemných prací bylo oznámkováno? [16]

Typové příklady standardní náročnosti

7. Z 10 studentů, mezi nimiž jsou Adam a Petr vybíráme tříčlennou komisi. Jaká je pravděpodobnost, že Adam nebo Petr budou mezi nimi? [$^{8}/_{15}$]
8. Hodíme 3-krát kostkou. Jaká je pravděpodobnost, že při 1. hodu nebo při 2. nebo při 3. padne sudé číslo? [$^{7}/_{8}$]
9. Každý ze spínačů je náhodně v poloze zapnuto nebo vypnuto nezávisle na druhých. Jaká je pravděpodobnost, že součástkou bude protékat el. proud?



   [$^{15}/_{64}$]

10. Žárovka svítí se spolehlivostí 85%. Jaká je spolehlivost systému (alespoň část svítí), jsou-li zapojeny: a) dvě žárovky sériově, b) dvě žárovky paralelně, c) dvě žárovky sériově a třetí k nim paralelně [0,723 ; 0,978 ; 0,958]
11. Test obsahuje 10 otázek, čtyři možné odpovědi, z nichž jedna je správná. Jaká je pravděpodobnost, že náhodným volením odpovědí vybereme alespoň 5 správných? [0,078]
12. Lék úspěšně léčí 90% případů onemocnění. Vypočítejte pravděpodobnost, že vyléčí alespoň 18 pacientů z 20, kterým je lék podán. [0,677]
13. Tabulka zaznamenává známky z matematiky pro 16 testů. Vypočítejte:
    a. Relativní četnost známky 2.
    b. Modus a medián známek z testu.
    c. Aritmetický průměr známek z testu.

Číslo testu	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Známka	1	5	2	3	5	4	3	4	2	2	4	4	5	2	3	4

[a) 25%, b) modus 4, medián 3,5, c) 53/16]

14. Průměrná cena 1 kg jablek je 27 Kč, průměrná cena 1 kg hrušek je 30 Kč. Nakoupili jsme jablka, hrušky a pomeranče. Čtvrtinu nakoupeného množství tvořily jablka, třetinu hrušky a zbytek tvořily pomeranče. Jaká je průměrná cena 1 kg pomerančů, jestliže byla průměrná cena 1 kg smíšeného zboží 32 Kč? [39,6 Kč]

Rozšiřující cvičení

15. **Státní maturita Matematika+ 2017** [A, N, A]

V tabulce jsou uvedeny výsledky soutěžících ve dvou různě početných skupinách A a B. Každý soutěžící mohl získat 0–4 body. Některé údaje v tabulce chybí, avšak víme:
V tabulce četností bude v každém řádku 5 různých čísel,
ve sloupcích bude vždy ve skupině B číslo o 2 větší než ve skupině A.

Skupina	Počet bodů	0	1	2	3	4	Aritmetický průměr	Medián	Modus
A	Četnost	2	0	8				2,5	4
B		4	2						

23 Po doplnění potřebných údajů rozhodněte o každém z následujících tvrzení (23.1–23.3), zda je pravdivé (A), či nikoli (N).

		A	N
23.1	U skupiny A je aritmetický průměr větší než medián.	☐	☐
23.2	U skupiny B je medián 2,5.	☐	☐
23.3	U skupiny B je aritmetický průměr 2,5.	☐	☐