

K problematice regresní analýzy

BOHUMIL VYBÍRAL

Přírodovědecká fakulta Univerzity Hradec Králové

1. Úvodní poznámka

Mezi středoškolskou mládeží orientovanou na přírodní a technické vědy se stává velmi populární regresní analýza. Umožňuje to její snadná dostupnost již na tzv. vědeckých kapesních kalkulátorech a na PC (např. v programu Excel). V případě regresní analýzy sice nejde o zcela novou metodu pro zpracování výsledků měření nebo různých empirických pozorování, avšak její praktické rozšíření umožnila až současná úroveň výpočetní techniky. Využívání této metody středoškolskou mládeží má však dosti podstatný nedostatek: mnohdy v podstatě neznají ani princip této metody, neboť její výklad patří až do aplikované matematiky na vysoké škole. Jak jsem si všiml při práci s talentovanou mládeží na různých soustředěních Fyzikální olympiády, studenti se při aplikaci této metody pak mohou dopouštět nežádoucích chyb nebo narážet na různá aplikační úskalí. Cílem této statě je podat stručný a srozumitelný výklad principu regresní analýzy a metodu demonstrovat na několika fyzikálních aplikacích. Článek volně navazuje na stať [3], publikovanou na stránkách tohoto časopisu.

2. Princip regresní analýzy

Regresní analýza je rigorózní analytická metoda prokládání křivek zpravidla jednoduchých funkčních závislostí empirickými body, které byly získány měřením závislostí fyzikálních jevů (nebo i jevů chemických, biologických, technických či společenských). Je založena na stejných metodách matematické statistiky, které užíváme při zpracování dat fyzikálních

měření. Její kořeny tedy sahají až ke Gaussovi. *Regrese*¹ neboli *regresní analýza* hledá vhodný statistický odhad (predikci) analytických závislostí mezi veličinami z empirických podkladů, získaných měření. Zpětně tedy hledáme regresní závislost např. sledovaných fyzikálních veličin. Cílem procesu regrese je nalezení (co nejlepší odhad) příslušné regresní funkce.

Mějme fyzikální veličinu y (např. elektrický odpor) a hledejme její závislost na nezávisle proměnné veličině x (např. na teplotě). Přímým měřením získáme n dvojic veličin $[x_1, y_1], [x_2, y_2], \dots, [x_n, y_n]$, které v kartézské soustavě os x, y můžeme znázornit jako *bodový graf*.

Předpokládejme, že mezi veličinami x, y existuje funkční vztah $y = f(x)$ známého tvaru (v případě naší sledované závislosti elektrického odporu např. lineární). Pokud by při uskutečňování měření nevznikaly náhodné chyby (víme, že to v principu není možné), pak by všechny body $[x_i, y_i]$, $i = 1, 2, \dots, n$, ležely na křivce $y = f(x)$. Ve skutečnosti však je $y_i = f(x_i) + \varepsilon_i$, kde ε_i je náhodná chyba i -tého měření. Vlivem chyb jsou tedy body $[x_i, y_i]$ rozptýleny kolem hledané regresní křivky, která má být obrazem funkce $y = f(x)$. Tato hledaná funkce obsahuje určitý počet neznámých konstant (parametrů) $b_0, b_1, \dots, b_p$. Můžeme tedy psát

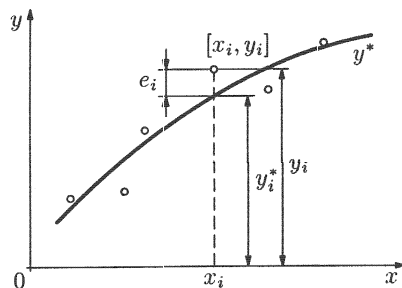
$$y = f(x; b_0, b_1, \dots, b_p). \quad (1)$$

Tyto neznámé parametry se nazývají *regresní koeficienty*. Např. lineární funkce $y = b_0 + b_1 x$ (resp. $y = a + bx$) obecně obsahuje dva neznámé regresní koeficienty. Měřením veličiny y tedy získáme hodnoty y_i , které můžeme vyjádřit rovnicí

$$y_i = f(x, b_0, b_1, \dots, b_p) + \varepsilon_i, \quad (i = 1, 2, \dots, n), \quad (2)$$

kde ε_i jsou zmíněné náhodné chyby měření. Máme-li body, které jsme získali měření, tj. $[x_1, y_1], [x_2, y_2], \dots, [x_n, y_n]$, proložit křivku (1), musíme provést statisticky odhad regresních koeficientů $b_0, b_1, \dots, b_p$. Požadavkem tohoto odhadu je, aby proložená křivka (1) „co nejlépe přiléhala“ empiricky získaným bodům (2). Regresní koeficienty b_j ($j = 0, 1, \dots, p$) získané statistickým odhadem označíme b_j^* . Nazývají se *výběrové regresní koeficienty*.

¹ *Regrese* je latinské slovo a znamená „zpětný pochod“.



Obr. 1 K principu regrese: $[x_i, y_i]$ empirický bod, y^* empirická regresní funkce, e_i reziduum měření

Výsledek odhadu regresních koeficientů závisí na tom, jaké kritérium „přiléhavosti“ regresní křivky k empirickým bodům zvolíme. K řešení využijeme *metodu nejmenších čtverců* známou z teorie chyb (např. [1], [4]). Empirickou regresní funkci označíme takto

$$y^* = f(x; b_0^*, b_1^*, \dots, b_p^*) \quad (3)$$

Nejprve definujeme veličinu *reziduum měření* jako rozdíl naměřené hodnoty y_i a hodnoty y_i^* vypočtené z empirické regresní funkce (3), tedy

$$e_i = y_i - y_i^* = y_i - f(x; b_0^*, b_1^*, \dots, b_p^*) \quad (4)$$

(viz obr. 1). Pomocí veličiny (4) definujeme *reziduální (zbytkový) součet čtverců*

$$S_e = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - y_i^*)^2 = \sum_{i=1}^n [y_i - f(x; b_0^*, b_1^*, \dots, b_p^*)]^2, \quad (5)$$

který se užívá ke statistickému odhadu regresních koeficientů b_j^* , $j \in \{0, p\}$. Podle metody nejmenších čtverců bude tento odhad nejlepší, když součet (5) nabude minima:

$$S_e = \min. \quad (6)$$

Nutnou podmínkou pro toto minimum je

$$\frac{\partial S_e}{\partial b_j^*} = 0 \quad \text{pro všechna } j \in \{0, 1, \dots, p\}. \quad (7)$$

Provedením příslušných $p+1$ derivací dostaneme soustavu $p+1$ lineárních rovnic o $p+1$ neznámých $b_0^*, b_1^*, \dots, b_p^*$, které se nazývají *normální rovnice*. Aby proces určení těchto neznámých měl smysl, musí pro počet n nezávislých měření být $n \geq p+1$. Pokud je $n = p+1$, prochází regresní křivka všemi empirickými body x_i, y_i , což v důsledku existence chyb měření není správné. Zpravidla tedy žádáme, aby

$$n > p+1. \quad (8)$$

Např. pro lineární regresní funkci $y = b_0 + b_1x$, kdy $p = 1$, to znamená $n > 2$; pro kvadratickou regresní funkci $n > 3$. Podmínka (8) znamená, že regresní křivka nemůže spojovat většinu bodů $[x_i, y_i]$ (anebo dokonce žádný), což je důsledek chyb měření.

Odhad regresních koeficientů přímým výpočtem z podmínek (7), tj. sestavení a řešení normálních rovnic, bývá úloha zdoluhavá a pro $p > 1$ náročná. V praxi se řeší pomocí počítače nebo (vědeckého) kalkulátoru – viz odst. 5. Nicméně z hlediska dobrého pochopení principu metody nejmenších čtverců a regresní analýzy vůbec, je užitečné ukázat a vyzkoušet si přímé řešení alespoň pro lineární regresní funkci, což je předmětem příkladu 2 v našem textu.

Vedle zde diskutovaného případu regrese funkce jedné proměnné, je rozpracována i regrese funkcí s několika nezávisle proměnnými (viz např. [2]). Je-li počet proměnných k , představuje x uspořádanou k -tici proměnných. Lze tedy x považovat za vektor o k složkách, který podmiňuje závisle proměnnou veličinu y . Těmito případy regrese se v našem textu dále nebudeme zabývat.

3. Typy regresních funkcí

V praxi se lze setkat s regresními funkcemi, které jsou uvedeny v tabulce 1.

I když funkce 1 a 2 jsou zvláštním případem funkce 3, je při regresní analýze vhodné postupovat specializovaně, jsou-li k tomu teoretické důvody. Je-li např. zřejmé, že lineární funkce musí procházet počátkem, zvolíme přímo modelovou funkci 2 (viz příklad 3). Pokud bychom v tomto případě zvolili obecný tvar přímky 3, regresní analýzou bychom pravděpodobně dospěli k odhadu $b_0^* \neq 0$, i když skutečná přímka musí procházet bodem $[0, 0]$.

Tab. 1 Přehled užívaných regresních funkcí

č.	Typ závislosti	Regresní funkce	Poznámka
1	konstantní	$y = b_0$	
2	lineární	$y = b_1 x$	přímka jde počátkem
3		$y = b_0 + b_1 x$	obecná přímka
4	kvadratická	$y = b_0 + b_1 x + b_2 x^2$	zvláštní případy polynomické závislosti
5	kubická	$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3$	
6	kvartická	$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + b_4 x^4$	
7	polynomická	$y = b_0 + b_1 x + \dots + b_r x^r$	
8	lineárně lomená	$y = a + \frac{b}{x}$	$x \neq 0$
9	exponenciální	$y = ab^x$	$b = e^B, B \neq 0, b > 0$
10		$y = a e^{Bx}$	$B = \ln b, b > 0$
11	logaritmická	$y = a + b \ln x$	$x > 0$
12	mocninná	$y = ax^b$	$a \neq 0$
13	logistická	$y = a + \frac{d}{1 + b e^{cx}}$	$b \neq 0, d \neq 0$
14	sinusoidální	$y = a + \sin(cx + d)$	

Počítačové a kalkulátorové programy nabízejí široké spektrum modelových funkcí, uvedených v tab. 1 – viz rovněž odst. 5. Přesto je někdy užitečné pomocí vhodné substituce *nelineární funkci* jednoduše *převést na lineární regresní funkci*. Řešení této úlohy je přehledně uvedeno v tabulce 2. Zde uvedená čísla nelineárních funkcí odpovídají číslům v tabulce 1.

Tab. 2 Převod některých nelineárních regresních funkcí na lineární funkce

	Nelineární regresní funkce	Substituce	Linearizované regresní funkce
8	$y = a + \frac{b}{x}$	$\frac{1}{x} = \xi$	$y = a + b\xi$
9	$y = ab^x$	$\ln y = \eta, \ln a = A, \ln b = B$	$\eta = A + Bx$
11	$y = a + b \ln x$	$\ln x = \xi$	$y = a + b\xi$
12	$y = ax^b$	$\ln y = \eta, \ln a = A, \ln x = \xi$	$\eta = A + b\xi$
	$y = a e^{\frac{B}{x}}$	$\ln y = \eta, \ln a = A, \frac{1}{x} = \xi$	$\eta = A + b\xi$

4. Hodnocení kvality modelu regrese

Kvalita zvoleného modelu regrese, tj. vhodnost volby určité regresní funkce a odhad jejich výběrových regresních koeficientů, se testuje. Jednou z výchozích veličin je *reziduální součet čtverců* S_e , definovaný vztahem (5), a to proto, že regresní koeficienty se odhadují právě tak, aby tento součet byl minimální – viz (6). Další pomocnou veličinou je *celkový (totální) součet čtverců*, definovaný vztahem

$$S_t = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n \left(y_i - \frac{1}{n} \sum_{i=1}^n y_i \right)^2. \quad (9)$$

K hodnocení kvality modelu regrese se užívají tyto veličiny:

1. *Koeficient determinace* r^2 modelu regrese definovaný vztahem

$$r^2 = 1 - \frac{S_e}{S_t}. \quad (10)$$

Zřejmě platí $0 \leq r^2 < 1$. Čím více se koeficient blíží k jedné, tím méně jsou body $[x_i, y_i]$ rozptýleny okolo regresní křivky (existují i výjimky z tohoto pravidla – viz níže uvedenou poznámku a rozbor v příkladu 4). Velikost $r^2 > 0,95$ se často považuje za dobré kritérium pro přijetí zvoleného modelu.

Poznámka. Hodnocení koeficientem r^2 má určité úskalí. Pro jednoduchou modelovou funkci $y = a^* = \text{konst.}$ je $S_t = S_e$ a tudíž $r^2 = 0$ bez ohledu na kvalitu modelu (viz příklad 4). Koeficient r^2 nelze tedy použít u funkce $y = a^* + b^*x$, je-li $b^* = 0$. Je-li $b^* \neq 0$, avšak velmi malé, je hodnocení pomocí r^2 rovněž málo vhodné, jak je ukázáno ve zmíněném příkladě 4.

2. *Koeficient korelace* r je druhou odmocninou koeficientu determinace (10). Užívá se u lineární regresní funkce.

3. *Reziduální rozptyl* s^2

$$s^2 = \frac{S_e}{n - (p + 1)}, \quad (11)$$

kde $p + 1$ je počet odhadovaných regresních koeficientů a $n - (p + 1) > 0$, tj. počet měření zmenšený o počet regresních koeficientů, se nazývá *počet stupňů volnosti* reziduálního součtu čtverců S_e .²

²Počet odhadovaných regresních koeficientů je roven $p + 1$ pouze v případě úplného polynomu stupně p .

4. *Směrodatná odchylka* s je druhou odmocninou rozptylu (11), tj.

$$s = \sqrt{\frac{S_e}{n - (p + 1)}}. \quad (12)$$

Má význam statistického odhadu směrodatné odchylky kteréhokoli měření y_i , tj. chyby měření. Pro hodnocení kvality modelu regrese má větší význam než koeficient determinace (10).

Pro jednotlivé regresní funkce lze odvodit specializované vztahy pro S_e , S_t a tedy i specializované vzorce pro veličiny (10), (12).

5. Praktikum regresní analýzy

Výpočty, které jsou spojeny s praktickým prováděním regresní analýzy, jsou složité a zejména zdoluhavé. Proto rozvoj aplikací regresní analýzy umožnila až současná úroveň výpočetní techniky. Programy pro provádění regresní analýzy jsou především součástí programového vybavení PC, např. program Excel od firmy Microsoft. Provádění regresní analýzy na PC vyniká komfortem – především je k dispozici jemná a rozměrná grafika a bezprostřední možnost tisku. Protože práce pomocí programu Excel bývá součástí školních cvičení v předmětu *Výpočetní technika*, nebudu se využitím a aplikací tohoto programu dále zabývat.

Existují však případy, kdy nemůžeme pohodově využít PC (např. při soutěži FO a na přípravných soustředěních FO). Pak se nabízí využití současných kapesních „vědeckých“ kalkulatorů, které bývají bohatě programově vybaveny i pro provádění regresní analýzy. Na současném trhu se jich vyskytuje celá řada (nejznámější jsou od výrobců Casio, Texas Instruments, Hewlett Packard, Sharp nebo Tronic). Např. kalkulatory Casio řady fx nabízejí modelové funkce 3, 4, 9, 11 a 12 z tab. 1. Grafické kalkulatory Texas Instruments TI 89/92 funkce 3, 4, 5, 6, 10, 11, 12, 13 a 14. Tyto grafické kalkulatory vyřeší nejen regresní koeficienty, nýbrž i směrodatné odchylky, koeficienty determinace (u vybraných funkcí) a umožní navíc zobrazit pomocí zvolených značek příslušné grafy, tj. bodový graf naměřených hodnot a proloženou regresní křivku. Tím lze získat okamžitý přehled o vyřešené regresní analýze problému. Podobně jsou vybaveny i grafické kapesní kalkulatory firmy Hewlett Packard.

Jak postupujeme při regresní analýze dat měření na kapesním kalkulatoru?

1. Nejprve musíme rozhodnout, jakou použijeme *modelovou regresní funkci*. Ověříme-li teoreticky známou funkční závislost fyzikálních veli-

čin, víme především, o jakou funkci jde anebo by mělo jít. Regresní analýzou poté konkretizujeme průběh funkce pro dané podmínky tím, že pravděpodobnostním výpočtem na kalkulátoru určujeme (odhadujeme) regresní koeficienty. Určením směrodatné odchylky (12) nebo koeficientu determinace (10) současně ověříme, jak se empirický průběh liší od teoretického. Stejně postupujeme při verifikaci hypotézy, kdy jistou závislost předpokládáme, např. podle provedené úvahy.

Pokud neznáme teoretický ani předpokládaný průběh zkoumané závislosti fyzikálních veličin, nakreslíme nejprve *bodový graf* empirických bodů $[x_i, y_i]$ a odhadneme vhodný typ funkce, která by mohla být proložena mezi empirickými body. Někdy nám pomůže vhodná substituce, např. $\xi = 1/x$, $\eta = 1/y$ nebo $\xi = \ln x$, $\eta = \ln y$. Vhodná někdy může být i jen jednostranná uvedená substituce, tj. jen pro x nebo pro y . U teplotních závislostí musíme často přepočíst teplotu t [°C] na jednotky v Kelvinech, tj. $T = (\{t\} + 273,15)$ K.

Můžeme se také rozhodnout jen pro lineární regresní funkci a provést příslušné substituce podle tabulky 2.

Jak je z uvedeného rozboru zřejmé, je volba vhodné modelové regresní funkce specificky závislá na zkoumané fyzikální zákonitosti. Vhodná (resp. nevhodná) volba však ovlivní výsledek celé regresní analýzy.

2. Kalkulátor přepneme do režimu regresních výpočtů, *vymažeme* statistické paměťové registry, nastavíme zvolenou modelovou regresní funkci a do vstupní paměti pečlivě vložíme všechna naše statistická data, tj. výsledky měření $[x_i, y_i]$ *ve správném pořadí* podle vzrůstající nebo klesající nezávisle proměnné.

3. V okamžiku otevření prvního ze statistických paměťových registrů, v nichž následně mají být uloženy výběrové regresní koeficienty, dojde automaticky ke spuštění statistického výpočtu. Jeho (téměř okamžitým) výsledkem jsou číselné hodnoty výběrových regresních koeficientů a ostatních výstupních statistických veličin, které se automaticky uloží do označených paměťových registrů. Získané výsledky vhodně zaokrouhlíme (na 2 až 3 platné cifry).

4. Pomocí koeficientu determinace (r^2) a směrodatné odchylky (s) posoudíme kvalitu regrese a zvolený model buď použijeme anebo zamítneme.

5. Vhodné je, pokud to (grafický) kalkulátor umožňuje, nakreslit ještě jak *bodový graf* měřených veličin, tak do něj také vypočtenou *regresní křivku*. Vizualně tak můžeme snadno posoudit korelaci empirických hodnot s vypočtenou regresní funkcí. Tím je proces regresní analýzy ukončen.

Příklad 1 – teplotní závislost elektrického odporu

Jsou dány výsledky měření elektrického odporu rezistoru v závislosti na teplotě podle tabulky 3.

Tab. 3 Data měření elektrického odporu

$t/^{\circ}\text{C}$	19,0	25,0	30,2	36,0	40,2	45,3	50,0
R/Ω	76,3	77,8	79,7	80,9	82,4	84,0	85,1

Proveďte regresní analýzu dat měření a určete: a) teplotní součinitel elektrického odporu při vztažené teplotě $t_0 = 0^{\circ}\text{C}$; b) odpor, který bude odpovídat teplotám $t_1 = 20,0^{\circ}\text{C}$ a $t_2 = 60,0^{\circ}\text{C}$; c) při jaké teplotě bude mít vodič odpor $R_3 = 82,0\ \Omega$.

Řešení

a) Z teorie předpokládáme, že modelová závislost elektrického odporu je lineární. Pak

$$R = a + bt = R_0(1 + \alpha t), \quad R_0 = a, \quad \alpha = b/a.$$

Provedením regresní analýzy dat z tab. 3 (v konkrétním případě byl použit kalkulátor Casio *fx-991W*) dostaneme: $a = R_0 = 70,8\ \Omega$, $b = 0,289\ \text{K}^{-1} \cdot \Omega$, teplotní součinitel elektrického odporu $\alpha = b/a = 4,08 \cdot 10^{-3}\ \text{K}^{-1}$. Koeficient korelace $r = 0,9985$, koeficient determinace $r^2 = 0,997$.

b) $t_1 = 20,0^{\circ}\text{C}$: $R_1 = 76,6\ \Omega$; $t_2 = 60,0^{\circ}\text{C}$: $R_2 = 88,1\ \Omega$.

c) $R_3 = 82,0\ \Omega$: $t_3 = 38,8^{\circ}\text{C}$.

Příklad 2 – odvození pro lineární regresní funkci

Pro získání lepší představy o metodě řešení regresní analýzou uvažujte jednoduchou modelovou situaci: lineární regresní funkci ve tvaru $y = bx$, tedy přímku jdoucí počátkem a proveďte příslušná odvození. Tedy užitím metody nejmenších čtverců odvoďte vztah pro odhad regresního koeficientu b^* , vztahu pro reziduální součet čtverců S_e a pro směrodatnou odchylku s , je-li dáno n empirických dat $[x_i, y_i]$, $i = 1, 2, \dots, n$ hledané funkční závislosti.

Řešení

Podle metody nejmenších čtverců musí reziduální součet čtverců být minimální, tj.

$$S_e = \sum_{i=1}^n (y_i - b^* x_i)^2 = \min.$$

Podmínka je splněna, když $\frac{\partial S_e}{\partial b_j^*} = 0$, tedy když (pro jednoduchost vynecháme meze sumace)

$$-2 \sum (y_i - b^* x_i) x_i = 0, \quad \text{z toho} \quad b^* = \frac{\sum x_i y_i}{\sum x_i^2}.$$

Reziduální součet čtverců pak je

$$\begin{aligned} S_e &= \sum (y_i - b^* x_i)^2 = \sum y_i^2 - b^* (2 \sum x_i y_i - b^* \sum x_i^2) = \\ &= \sum y_i^2 - \frac{(\sum x_i y_i)^2}{\sum x_i^2}. \end{aligned}$$

Směrodatnou odchylku kteréhokoli měření vypočteme ze vztahu (12), kde v našem případě počet stupňů volnosti je $n - 1$. Tedy

$$s = \sqrt{\frac{S_e}{n - 1}}.$$

Směrodatná odchylka regresního koeficientu b^* je (viz [2])

$$s_{b^*} = \frac{s}{\sqrt{\sum x_i^2}}. \quad (13)$$

Příklad 3 – tíhové zrychlení z volného pádu

Při volném pádu olověné kuličky byly pro zvolené úseky dráhy s naměřeny doby t , které jsou uvedeny v tabulce 4. Naměřené hodnoty můžeme doplnit ještě o údaj $t = 0$ pro $s = 0$.

Tab. 4 Data měření volného pádu

s/m	4,00	5,20	6,00	7,00	8,00	9,00	10,00	11,00
t/s	0,90	1,03	1,10	1,20	1,28	1,35	1,43	1,50

Řešení

Protože kulička má velkou hustotu, můžeme pro relativně malé rychlosti zanedbat odpor vzduchu a pro závislost dráhy na čase brát vztah

$s = gt^2/2$. Zavedením substituce $t^2 = 2x$ a označením $s = y$, $g = b$ můžeme závislost vyjádřit lineární modelovou regresní funkcí $y = bx$. Odhad regresního koeficientu b^* (tj. tíhového zrychlení g) a jeho směrodatnou odchylku určíme z dat v tabulce 4 užitím vzorců odvozených v příkladu 2. Výpočty jsou uvedeny v tabulce 5. Výsledek řešení tíhového zrychlení je dobře přijatelný.

Tab. 5 Výpočet tíhového zrychlení regresní analýzou

t_i/s	x_i/s^2	y_i/m	$\sum x_i^2 = 5,14232 \text{ s}^4, \sum y_i^2 = 494,04 \text{ m}^2$ $\sum x_i y_i = 50,40269 \text{ m} \cdot \text{s}^{-2}$ $b^* = g = \frac{50,40269}{5,14232} \text{ m} \cdot \text{s}^{-2} = 9,8015 \text{ m} \cdot \text{s}^{-2}$ $S_e = \left(494,04 - \frac{50,40269^2}{5,14232} \right) \text{ m}^2 = 0,01568 \text{ m}^2$ $s = \sqrt{\frac{0,01568}{8}} \text{ m} = 0,0443 \text{ m} \doteq 0,05 \text{ m}$ $s_{b*} = \frac{0,0443}{\sqrt{5,142}} \text{ m} \cdot \text{s}^{-2} \doteq 0,02 \text{ m} \cdot \text{s}^{-2}$
0	0	0	
0,90	0,405	4,0	
1,03	0,53045	5,2	
1,10	0,605	6,0	
1,20	0,720	7,0	
1,28	0,8192	8,0	
1,35	0,91125	9,0	
1,43	1,02245	10,0	
1,50	1,125	11,0	
$g = (9,80 \pm 0,02) \text{ m} \cdot \text{s}^{-2}$			

Příklad 4 – regresní analýza dvou blízkých souborů

Uvažujme dva „cvičné“ blízké soubory A, B proměnných $[x_i, y_i]$ podle tabulky 6, které se vyznačují tím, že střední hodnoty $\bar{y}_A = \bar{y}_B = 10,0$. Proveďte regresní analýzu dat těchto souborů, tj. proveďte odhad regresních koeficientů, určete celkový součet čtverců (S_t), reziduální součet čtverců (S_e), koeficient determinace (r^2) a směrodatnou odchylku (s). Proveďte diskusi výsledků pro oba soubory.

Tab. 6 Číselné hodnoty souborů A, B

x_i		1	2	3	4	5	6	7	8	$\bar{y}$
A	y_i	10,0	10,1	10,0	9,9	10,0	9,9	10,0	10,1	10,0
B	y_i	9,9	9,9	10,0	10,0	10,0	10,0	10,1	10,1	10,0

Řešení Pro oba soubory lze použít modelovou lineární regresní funkci ve tvaru $y = a + bx$.

Soubor A

Použitím kalkulátoru dostaneme odhad koeficientů: $a^* = \bar{y} = 10,0$; $b^* = 0$ a součty čtverců:

celkový: $S_t = \sum (y_i - \bar{y})^2 = 0,04$ a reziduální: $S_e = \sum (y_i - y_i^*)^2 = 0,04$.

Koeficient determinace $r^2 = 1 - S_e/S_t = 0$,

směrodatná odchylka $s = \sqrt{\frac{S_e}{n - (p + 1)}} = \sqrt{\frac{0,04}{7}} = 0,0756 \doteq 0,08$.

Soubor B

V tomto souboru mají data stejnou velikost (avšak poněkud jiné uspořádání). Užijeme stejnou regresní funkci jako pro soubor A. Výpočtem na kalkulátoru zde dostaneme pro odhad regresních koeficientů jiné hodnoty: $a^* = 9,871$; $b^* = 0,02857$. Určíme nyní odhad jednotlivých hodnot y_i z regresní funkce $y_i^* = a^* + b^*x_i$; výsledky výpočtů jsou v tabulce 7.

Tab. 7 Výpočty pro soubor B

x_i	$y_i^* = a^* + b^*x_i$	
1	9,8996	$S_e = \sum (y_i - y_i^*)^2 = 0,0005713$
2	9,9281	
3	9,9567	
4	9,9853	$S_t = \sum (y_i - \bar{y})^2 = 0,040000$
5	10,0139	$r^2 = 1 - S_e/S_t = 0,8571$
6	10,0424	$s = \sqrt{\frac{S_e}{n - (p + 1)}} = \sqrt{\frac{0,005713}{6}} = 0,0309 \doteq 0,03$
7	10,0710	
8	10,0996	
		Kontrolní výpočet: $\bar{y}^* = 9,999575 \doteq 10,0$

Diskuse výsledků

- Pro soubor A vyšel koeficient determinace $r^2 = 0$, i když použitá modelová regresní funkce $y^* = a^* = \text{konst.}$ je bezpochyby vhodná. Je totiž zřejmé, že odchylky jednotlivých dat od y^* (tedy rezidua měření) nepřesahují $\pm 1 \%$. Protože v případě této regresní funkce je $S_t = S_e$, vychází zde z definice $r^2 = 0$ a veličina r^2 se tedy pro hodnocení kvality modelu regrese nehodí (obecně platí, že pro $b^* = 0$ je vždy $r^2 = 0$).

2. U druhého souboru B (oproti souboru A se liší jen pouhým přeskupením několika čísel) jsou hodnoty neklesající a proto $b^* \neq 0$. I tak je $r^2 = 0,8571 < 0,95$ a koeficient determinace tedy hodnotí použitý model jako málo vhodný, i když vhodnější model lze najít jen obtížně (největší reziduum je pro x_3 : $e_3 = +0,43 \%$ a pro x_6 : $e_6 = -0,43 \%$). Hodnocení podle s je příznivější než u modelové funkce souboru A.
3. Použijeme-li pro soubor B přímo modelovou funkci $y = a$, dává regresní analýza stejné výsledky jako pro soubor A.

L i t e r a t u r a

- [1] *Brož, J. at al.*: Základy fyzikálních měření, I. díl. SNP, Praha 1967.
- [2] *Rektorys, K. at al.*: Přehled užití matematiky. SNTL, Praha 1963; 6. vydání: Prometheus, Praha 1995.
- [3] *Vybíral, B.*: Jak správně a rychle zpracovat fyzikální měření. Matematika, fyzika, informatika. 13 (5/2004), s. 279–288, ISSN 1210–1761.
- [4] *Vybíral, B.*: Zpracování dat fyzikálních měření. Knihovnička fyzikální olympiády č. 52. MAFY, Hradec Králové 2002. ISBN80-86148-54-8.

Hrajeme si s fyzikou aneb jednoduché pokusy pro malé i velké žáky

2. část – KAPALINY

JANA ČESÁKOVÁ – MICHAELA KŘÍŽOVÁ

Přírodovědecká fakulta UHK, Hradec Králové

Tento článek je pokračováním článku o seminářích zaměřených na jednoduché fyzikální pokusy, které ve spolupráci s Univerzitou Hradec Králové pravidelně organizuje Školské zařízení pro další vzdělávání pedagogických pracovníků Královéhradeckého kraje [1]. V první části jsme představili pokusy na téma vzduch. Nyní bychom vám rádi přiblížili to, co jsme vyzkoušeli s učiteli na druhém semináři. Tématem byly KAPALINY.

Semináře pro učitele

Pro učitele fyziky připravujeme semináře s názvem „Hrajeme si s fyzikou aneb jednoduché pokusy pro malé i velké žáky“. Zde s učiteli procházíme mnoho zajímavých pokusů, které najdete i na akcích Hrajme si i hlavou [2] či Veletrh nápadů učitelů fyziky [3], řešíme různá vylepšení a varianty. Během čtyř seminářů jsme se zatím dotkli témat – voda a vzduch, fyzikální kouzla, zajímavé kapaliny či zvuk.

Druhý seminář byl zaměřen na KAPALINY. Hráli jsme si s kapalným dusíkem a newtonovskými kapalinami jako jsou např. inteligentní modelíny nebo směsi vody a škrobu. Zajímavostí byla i výroba hřejivých sáčků. Předvedli jsme i mnoho dalších pokusů např. jak sebrat olej z hladiny, jak udělat mrak či tornádo v lahvi a nakonec jsme spočítali, kolik špendlíků se vejde do skleničky plné vody. Pokusy tedy jednak pomáhaly s tématy hydrostatický i atmosférický tlak, Archimédův zákon, povrchové napětí, hustota či změny skupenství a dále motivovaly děti přemýšlet a zajímat se o věci kolem sebe.

V tomto příspěvku představíme malou ochutnávku z pokusů s kapalným dusíkem, inspirovaných experimenty, které provádí *doc. RNDr. Miloš Rotter, CSc.* z MFF UK [5].

„Hrnečku, vař“

Kapalný dusík není z těch kapalin, se kterými se běžně setkáváme. Proto o pokusy s ním byl velký zájem na již zmiňovaných akcích Hrajme si i hlavou.

Kapalný dusík jsme pro naše účely koupili od firmy Linde Gas a. s., která se zabývá výrobou a prodejem různých technických plynů. Cena jednoho litru kapalného dusíku není přemrštěná, pohybuje se v řádech desítek korun. Jen je trochu problém s jeho uchováváním. Pro naše pokusy nám stačilo několik starých termosek na zmrzlinu. Je nutné mít u nich provrtnané víčko, nebo víčko jen lehce posadit na otvor termosky. Z jednoho litru kapaliny se totiž za normálního tlaku a teploty 20 °C odpaří přibližně 680 litrů plynu, který musí mít možnost někam unikat. Proto nikdy nezavírejte termosku!!!

Kapalný dusík není nebezpečný ani zdravotně škodlivý, pokud se s ním manipuluje opatrně, aby nedošlo k popálení pokožky či výbuchu díky skladování v uzavřeném prostoru. Dávejte pozor, aby vaše pokožka byla v kontaktu s kapalným dusíkem co nejkratší dobu a na pokusy si vždy sundejte prstýnky. Dětem můžete ukázat jednoduché pokusy s kapalným dusíkem,

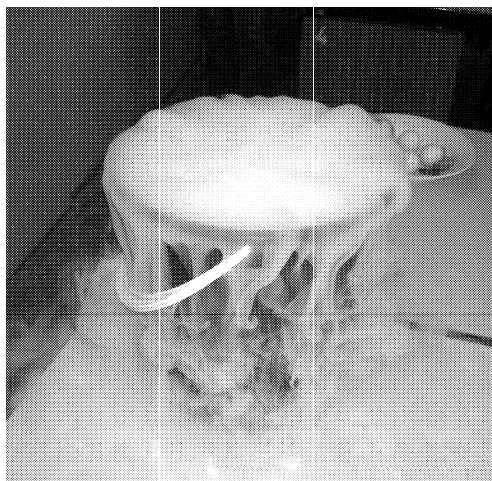
kteře jsou ale velmi efektní a děti nadchnou. Zde uvádíme dva návody na jejich realizaci. Na pokusy budete kromě kapalného dusíku potřebovat běžně dostupné věci – kbelík nebo umyvadlo s vodou, polystyrénovou vaničku (do které budete lít kapalný dusík a dávat do něj postupně předměty, které chcete ochladit), prkénko, kladívko, žvýkačku nebo gumu, kousek pomerančové kůry, růži a kleště (na vyjmutí ochlazených předmětů z kapalného dusíku).

Pokus „hrnečku vař“ je velmi jednoduchý. Stačí nalít kapalný dusík do kbelíku nebo umyvadla s vodou. Chladný, odpařující se dusík se tak smíchá s vodními parami, které kondenzují do mikroskopických kapiček a vytvoří hustou mlhu (obr. 1). Velmi efektní variantou pokusu je, nechat do umyvadla stále přitékat vodu z kohoutku. „Hrneček“ tak bude „vařit“ pěkně dlouho.

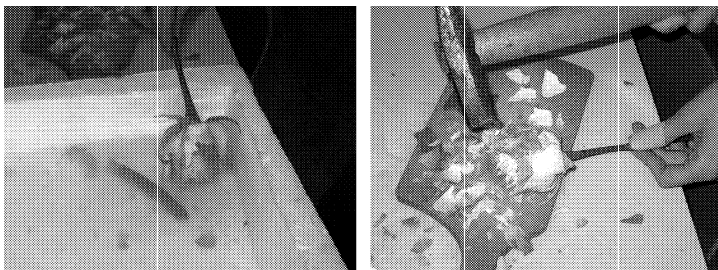
Pro zvýšení zájmu žáků můžeme použít videa z internetového portálu youtube.com, kterých najdete celou spoustu. Tento pokus je totiž oblíbený ve větší variantě – místo kbelíku lze použít bazén.

Jak kladívkem rozbít růži

Kapalný dusík je velmi „studený“. Jeho teplota varu je za normálního atmosférického tlaku $-195,8\text{ }^{\circ}\text{C}$. Když kapalným dusíkem ochladíme nějaký kov nebo plast, změní se jejich mechanické vlastnosti a stanou se



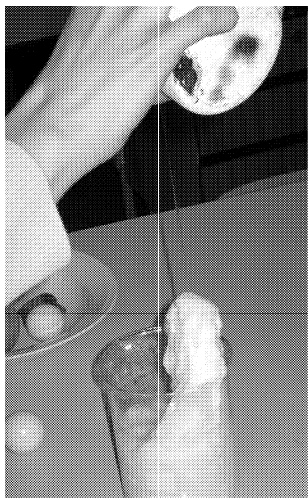
Obr. 1



Obr. 2

křehkými. To můžeme dětem nejlépe ukázat na balonku, gumě nebo žvýkačce. Po ochlazení je můžeme rozbít kladívkem na malé kousky!

Zkřehnutí můžeme demonstrovat také na nějakém předmětu rostlinného původu, například na slupce pomeranče nebo růži. Ochlazenou slupku pomeranče rozbijeme kladívkem díky tomu, že obsahuje vodu, která se ochlazením změní na led. Efektním pokusem je rozbítí růže zmražené v kapalném dusíku.



Obr. 3

do roztoku nespadne kousek námrazy). Po zchladnutí stačí do vody sáhout prstem nebo špejlí nebo začít vylévat, jako je vidět na obr. 3. Roztok

„Domácí“ hřejivé sáčky

Hřejivé sáčky patří ke známým zdravotním potřebám. Jsou plněny přesyceným vodným roztokem octanu sodného a nesmí obsahovat krystalizační jádra. Ta se vytvoří až klepnutím na kovový plíšek umístěný uvnitř sáčku, čímž dojde ke krystalizaci a uvolnění krystalizačního tepla. Obsah sáčku se zahřeje na teplotu až $55\text{ }^{\circ}\text{C}$.

Pomůcky: octan sodný, voda, hořák, mrazák

Postup: Na mírném plameni vytvoříme přesycený vodný roztok octanu sodného tak, aby neobsahoval žádná krystalizační jádra – jednoduše nesmí docházet k samovolné krystalizaci. Roztok dáme do mrazáku a necháme vychladit (doporučujeme něčím přikrýt, aby

byl ve stavu podchlazené kapaliny, což znamená, že měl nižší teplotu, než je jeho teplota tuhnutí, a přesto neztuhl. Dotekem vytvoříme krystalizační jádro, kolem kterého se molekuly začnou rychle seskupovat a uvolní při této krystalizaci teplo (jedná se o exotermní reakci).

Hřejivé sáčky lze i po proběhlé krystalizaci a ochlazení znovu použít. Stačí sáček zahřívat ve vroucí vodě tak dlouho, až obsahuje jen tekutinu. Pak ho můžeme opět použít, až budeme potřebovat ohřát ruce. Stejně tak i namíchaný roztok po ohřátí bude znovu fungovat.

Určitě také stojí za to prolistovat knížku [6], kde najdete návody na zajímavé pokusy ze světa tepla, jeho vlastností, šíření, projevů i jeho měření. Najdete zde další informace nejen o hřejivých polštářcích, ale např. i o mrazících směsích.

Směs vody a škrobu

V poslední době se již několikrát hovořilo o zajímavé směsi vody a škrobu, která fyzikálně spadá pod nenewtonovské kapaliny neboli kapaliny, které se chovají nějak podivně. Tato směs se chová tak, že při rychlých pohybech tuhne a stává se z ní téměř pevná látka. Naopak při pomalých pohybech se tváří jako kapalina. Tyto vlastnosti jsou způsobeny zajímavým tvarem molekul škrobu. Za normálních okolností volně kloužou po sobě a po molekulách vody, ale při prudkém pohybu se semknou k sobě a vypudí většinu vody pryč. Proto když do takovéto směsi bouchneme pěstí nebo kladívkem, rozhodně nám nevycákne a ani nepronikneme pod hladinu. Dokonce může hladina popraskat jako vysušená jílovitá zem. Ale na druhou stranu zcela bez problémů můžeme pod hladinu pomalu zasunout celou ruku. Podobně uhněteme z kapaliny kuličku, která se po uvolnění celá rozteče.

K podobným pokusům potřebujete jen škrob (nám se více osvědčil kuříčkový, ale funguje i bramborový) a vodu. Směs se dá namíchat bez použití speciálních vzorců. Trocha škrobu a trocha vody a správný poměr doladíme případným dosypáním škrobu či dolitím vody.

Pokud budete mít dostatek škrobu, můžete si vyzkoušet, jaké je to chodit po hladině a nořit se do bažiny. Abyste se však neponořili, musíte se pohybovat rychle. Jak vidíte na fotografiích (obr. 4), baví to velké i malé.

Závěr

V posledních letech jsou populární fyzikální dárky – zvláštnůstky, které fungují podle fyzikálních zákonů, v moderním provedení. Patří mezi ně in-

teligentní plastelína, která vám za několik hodin proteče malým otvorem, a přitom ji rozbijete kladívkem, jako by byla z porcelánu. Na seminářích jsme ukazovali plastelíny magnetické nebo teplocitlivé. Zajímavý je také hydrofobní písek, který odpuzuje vodu, a můžete s ním ukazovat, proč na listech rostlin vznikají kapičky rosy. A to je opravdu jen zlomek toho, co se dá pořídit. Pokud vám školní rozpočet nedovolí podobné hračky kupovat a ani nemáte ve třídě žáky, kteří by hračky nosili do školy na hraní, můžete je alespoň vzít jako inspiraci a snažit se je převést do jednodušších a levnějších pomůcek. O to větší potom vzbudí zájem, když je žáci budou znát nebo je budete moci ukázat, v jiné podobě. Všechny podivné novinky a pokusy dělají fyziku pro děti zajímavější a díky tomu třeba najdou kouzlo, které v sobě fyzika bezesporu má!



Obr. 4

Literatura

- [1] <http://www.cvkhhk.cz/>
- [2] <http://www.hrajme-si-i-hlavou.cz/>
- [3] <http://kdf.mff.cuni.cz/veletrh/sbornik>
- [4] <http://www.youtube.com/>
- [5] http://kdf.mff.cuni.cz/veletrh/sbornik/Veletrh_07/07_19_Rotter.html
- [6] Polák, Z.: Hrátky s teplem, Praha, ČEZ, 2007.