Úloha V.S ... lineární

10 bodů; průměr 8,38; řešilo 21 studentů

a) *Zkuste vlastními slovy popsat, k čemu a jak se používá lineární regrese (postačí vlastními slovy popsat následující: dva hlavní případy aplikace lineární regrese, používaný model, předpoklady modelu, postup volby prokládané funkce, způsob vyjádření nejistot měření, základní grafické metody regresní diagnostiky). Není potřeba uvádět přesná matematická odvození, stačí požadované pojmy a vlastnosti stručně popsat.*

b) *V přiloženém datovém souboru* linreg1.csv *naleznete výsledky určitého fyzikálního experimentu, ve kterém jsme měřili dvojice dat (x_i, y_i). Naměřenými daty chceme proložit teoretickou funkci, kterou je v tomto případě parabola, tedy funkce tvaru*

$$f(x) = ax^2 + bx + c\,.$$

Hlavním cílem experimentu je určit hodnotu koeficientu a (tedy koeficient u x^2). Určete hodnotu tohoto koeficientu včetně nejistoty měření. Není potřeba provádět regresní diagnostiku.

c) *V přiloženém datovém souboru* linreg2.csv *naleznete výsledky určitého fyzikálního experimentu, ve kterém jsme měřili dvojice dat (x_i, y_i). Naměřenými daty chceme proložit teoretickou funkci, kterou je v tomto případě logaritmická funkce, tedy funkce tvaru*

$$f(x) = a + b \cdot \log(x)\,.$$

Hlavním cílem zpracování dat je vykreslit graf naměřených dat spolu s proloženou teoretickou závislostí. Vykreslete takovýto graf (včetně intervalového odhadu pro prokládanou funkci) a stručně ho okomentujte (takovýto graf musí mít všechny náležitosti). Není potřeba provádět regresní diagnostiku.

d) *Předpokládejme, že máme naměřeny dvojice dat (x_i, y_i) a chceme jimi proložit lineární funkční závislost, tedy funkci tvaru*

$$f(x) = a + bx\,.$$

Odvoďte přesnou podobu vzorce na výpočet hodnoty odhadů regresních koeficientů. Můžete použít libovolnou ze dvou metod představených v seriálu a také libovolné jiné zdroje, pokud je budete řádně citovat. Vzorec chceme opravdu odvodit (tj. uvést výpočet), nikoliv pouze napsat.

Bonus: *V úlohách* b) a c) *proveďte regresní diagnostiku a diskutujte, zda jsou splněny všechny potřebné předpoklady (pokud to jde, proveďte také test vhodnosti prokládané funkce a diskutujte jeho výsledky).*

Pro práci s daty použijte výpočetní prostředí R. *Pro vyřešení těchto úkolů postačí drobně upravit přiložený skript, ve kterém je pomocí komentářů v kódu vysvětlena potřebná syntaxe jazyka* R.

Michal někde slyšel, že lineární regrese je prý úplně jednoduchá věc.

a) Detailní odpověď na tuto otázku dostaneme jen přečtením pátého dílu seriálu. Na tomto místě popíšeme jen opravdu nejzákladnější věci.

Lineární regrese se může použít v případech, kdy máme naměřená data ve formě dvojic $(x_1, y_1), \ldots, (x_n, y_n)$ a chceme jimi proložit teoretickou funkční závislost, která je lineární v neznámých regresních koeficientech (více o používaném modelu v dalším odstavci). Dva

hlavní důvody, proč používat lineární regresi jsou, že chceme nepřímo změřit nějakou fyzikální veličinu (tedy hodnotu vybraného regresního koeficientu) nebo že chceme vykreslit graf s naměřenými daty, ve kterém bude pro názornost vykreslena teoretická funkční závislost.
Nyní si blíže popíšeme model, který při aplikaci lineární regrese pro naše data předpokládáme. Lineární regresní model předpokládá, že měřená data splňují vztah

$$y_i = \beta_0 + \beta_1 f_1(x_i) + \dots + \beta_k f_k(x_i) + \varepsilon_i \,,$$

kde $\beta_0, \beta_1, \dots, \beta_k$ jsou neznámé regresní koeficienty, které chceme odhadovat pomocí naměřených dat, $f_1(x), \dots, f_k(x)$ jsou zvolené prokládané funkce a ε_i je náhodná nepřesnost měření (tedy náhodná veličina).
Lineární regresní model má 4 předpoklady, kterými jsou

- Správná volba prokládané funkce.
- Stejný rozptyl pro všechna měření.
- Nezávislost jednotlivých měření.
- Normální rozdělení našich měření.

Fyzik by se měl vždy snažit při měření dat zajistit splnění těchto předpokladů (hlavně předpoklad o nezávislosti jednotlivých měření) a po použití metod lineární regrese zkontrolovat, zda byly všechny potřebné předpoklady splněny (o způsobu ověřování splnění předpokladů v dalších odstavcích).
Abychom minimalizovali riziko špatné volby prokládané funkce, je dobré volit jen takové funkce, které mají určité teoretické fyzikální opodstatnění. Pokud toto nebudeme dodržovat, vystavujeme se nebezpečí, že prokládanou funkci zvolíme špatně a všechny naše výsledky budou neplatné.
Podobně jako při měření jedné fyzikální veličiny můžeme i v případě lineární regrese vyjádřit odhady regresních koeficientů a prokládané funkce včetně nejistot měření. V textu seriálu byly uvedeny přesné vzorce na výpočet těchto nejistot, které zde nebudeme opakovat. V praxi je za nás bude vždy numericky počítat matematický software.
Jak bylo řečeno v předchozích odstavcích, vždy by se měla ověřit platnost předpokladů použitého modelu lineární regrese. K tomu se používají residua modelu definovaná jako

$$u_i = y_i - \widehat{\beta_0} - \widehat{\beta_1} f_1(x_i) - \dots - \widehat{\beta_k} f_k(x_i) \,,$$

kde $\widehat{\beta_0}, \dots, \widehat{\beta_k}$ jsou odhadnuté hodnoty regresních koeficientů. Pokud byly všechny předpoklady použitého lineárního regresního modelu splněny, měla by být residua navzájem nezávislá a neměla by vykazovat žádnou (ani lokální) závislost na nezávisle proměnné. Toto se může následně ověřovat na vhodných grafech residuí (např. residua oproti nezávisle proměnné, residua oproti hodnotám proložené funkce, residua oproti posunutým residuím atd.).

b) Ve výpočetním prostření *R* načteme poskytnutý datový soubor na pomocí lineární regrese naměřenými daty proložíme funkci tvaru

$$f(x) = ax^2 + bx + c \,.$$

Výpočetní prostředí *R* spočte podle vzorců uvedených v textu seriálu pro zadaný model odhady regresních parametrů včetně nejistot měření a uvede je na výstupu. Pro tato konkrétní data a prokládanou funkci dostaneme odhad regresního koeficientu u x^2

$$\widehat{a} \doteq 1{,}98$$

a nejistotu měření tohoto koeficientu

$$s_a \doteq 0{,}10\,.$$

Když použijeme zkrácený zápis, můžeme psát, že jsme nepřímým měřením určili hodnotu koeficientu a jako

$$a = (1{,}98 \pm 0{,}10)\ .$$

c) Ve výpočetním prostředí R proložíme poskytnutými daty funkci tvaru

$$f(x) = a + b \cdot \log(x)\,.$$

Výpočetní prostředí R spočítá odhady regresních koeficientů včetně nejistot měření podle vzorců uvedených v textu seriálu. Pro naše konkrétní data dostaneme hodnoty

$$\begin{aligned} a &= (5{,}0 \pm 0{,}08)\,,\\ b &= (2{,}87 \pm 0{,}09)\,, \end{aligned}$$

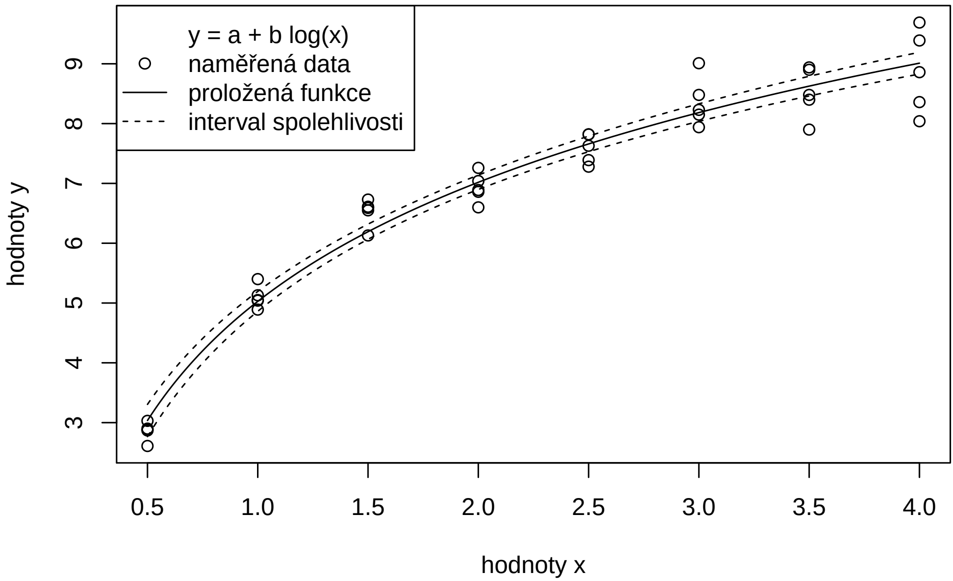
kde využíváme klasické zkrácený zápis nejistot měření. Na Obr. 1 můžeme vidět měření včetně proložené funkce a intervalu spolehlivosti pro prokládanou funkci s hladinou spolehlivosti 95%. Interval spolehlivosti je poměrně úzký, což značí, že jsme pomocí naměřených dat určili prokládanou funkci poměrně přesně. Nemělo by nás nijak překvapit, že mnoho naměřených bodů (většina) padne mimo interval spolehlivosti pro prokládanou funkci. Toto je způsobeno tím, že když zkombinujeme informaci ze všech měření pomocí lineární regrese, jsme schopni určit hodnoty prokládané funkce velice přesně.
Z tohoto obrázku můžeme vidět, že proložená funkce na naměřená data sedí poměrně dobře, ale je možné, že je mírně porušen předpoklad homoscedasticity (vidíme, že rozptyl měřených hodnot závisí na nezávsle proměnné). Ověřování splnění všech předpokladů se budeme podrobněji věnovat v bonusové části tohoto vzorového řešení.

d) V textu seriálu byly zmíněny celkem 2 metody, jak se dají vyjádřit odhady regresních parametrů v lineární regresi pomocí naměřených dat. Ve vzorovém řešení pro jistotu uvedeme oba možné postupy.
i Jako první vyjádříme tyto odhady nalezením minima součtu čtverců pomocí jejich parciálních derivací, které položíme rovny nule a vyřešíme příslušné rovnice (obecný postup hledání extrému funkcí více proměnných). Nejprve si musíme vyjádřit součet čtverců

$$S(a,b) = \sum_{i=1}^{n} \left(y_i - a - bx_i\right)^2\ .$$

Tento součet čtverců nyní parciálně zderivujeme podle obou proměnných a a b a tyto parciální derivace položíme rovny 0, čímž dostaneme

$$\begin{aligned} \frac{\partial S(a,b)}{\partial a} &= \sum_{i=1}^{n} -2\left(y_i - a - bx_i\right) = 0\,,\\ \frac{\partial S(a,b)}{\partial b} &= \sum_{i=1}^{n} -2x_i\left(y_i - a - bx_i\right) = 0\,. \end{aligned}$$



Obr. 1: Graf s naměřenými daty, proloženou funkční závislostí a 95% intervalem spolehlivosti pro prokládanou funkci.

Nyní musíme nalézt řešení této soustavy rovnic. Úpravami těchto dvou rovnic dostáváme

$$\sum_{i=1}^{n} y_i - na - b\sum_{i=1}^{n} x_i = 0\,,$$

$$\sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2 = 0\,.$$

Z první rovnice nyní dostáváme, že platí

$$a = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i\right)\,.$$

Toto vyjádření můžeme nyní dosadit do druhé rovnice a dostáváme

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\left(\frac{1}{n}\sum_{i=1}^{n} y_i - b\frac{1}{n}\sum_{i=1}^{n} x_i\right) - b\sum_{i=1}^{n} x_i^2 = 0\,.$$

Z této rovnice už se jednoduchými algebraickými úpravami dá odvodit vzorec pro b, který má tvar

$$\widehat{b} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i \right)^2} .$$

Zpětným dosazením potom získáme vzorec pro odhad parametru a ve tvaru

$$\widehat{a} = \frac{\frac{1}{n} \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i \right)^2} .$$

Tímto jsme tedy odvodili vzorce na výpočet regresních koeficientů pomocí minimalizace součtu čtverců residuí v našem lineárním regresním modelu.
Nyní si v rychlosti uvedeme, jak by se dalo postupovat druhou alternativní cestou, kterou jsme zmínili v textu seriálu. Nejprve si musíme uvědomit, jak vypadá matice modelu a vektor naměřených hodnot. V našem konkrétním případě budou mít tvar

$$\mathbb{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} , \quad \mathbf{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} .$$

Nyní stačí tuto matici a vektor dosadit do vzorce na výpočet odhadů regresních koeficientů

$$\begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \mathbf{Y} ,$$

Pro snadnější výpočet budeme postupně vyčíslovat uvedené matice. Nejprve se budeme zabývat maticí $\mathbb{X}^T \mathbb{X}$, vyčíslíme tuto matici

$$\mathbb{X}^T \mathbb{X} = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix} .$$

Nyní si musíme spočítat inverzní matici $\left(\mathbb{X}^T \mathbb{X} \right)^{-1}$. Budeme postupovat klasickým způsobem, že si vedle této matice napíšeme jednotkovou matici a budeme dělat stejné řádkové úpravy na obě matice až dostaneme místo původní matice jednotkovou matici. Matice, která se potom objeví na místě původní jednotkové matice bude hledaná inverzní matice.

$$\left(\begin{array}{cc|cc} n & \sum_{i=1}^{n} x_i & 1 & 0 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & 0 & 1 \end{array} \right) \sim \left(\begin{array}{cc|cc} n & \sum_{i=1}^{n} x_i & 1 & 0 \\ 0 & \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i \right)^2 & -\frac{1}{n} \sum_{i=1}^{n} x_i & 1 \end{array} \right) \sim$$

$$\sim\left(\begin{array}{cc|cc} n & 0 & 1+\frac{\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2}{\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2} & -\frac{\sum\limits_{i=1}^{n}x_i}{\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2} \\ 0 & 1 & \frac{-\frac{1}{n}\sum\limits_{i=1}^{n}x_i}{\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2} & \frac{1}{\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2} \end{array}\right)\sim$$

$$\sim\left(\begin{array}{cc|cc} 1 & 0 & \frac{\frac{1}{n}\sum\limits_{i=1}^{n}x_i^2}{\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2} & -\frac{\frac{1}{n}\sum\limits_{i=1}^{n}x_i}{\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2} \\ 0 & 1 & \frac{-\frac{1}{n}\sum\limits_{i=1}^{n}x_i}{\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2} & \frac{1}{\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2} \end{array}\right).$$

V průběhu těchto úprav jsme používali jen odečítání násobku jednoho řádku od druhého a násobení řádků konstantou. Konkrétně jsme v první úpravě odečetli $\frac{1}{n}\sum\limits_{i=1}^{n}x_i$násobek prvního řádku od druhého. Ve druhé úpravě jsme druhý řádek vydělili $\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2$, následně jsme odečetli $\sum\limits_{i=1}^{n}x_i$násobek tohoto řádku od prvního řádku. Ve třetí úpravě jsme pouze vydělili první řádek číslem n a upravili zlomek do hezčího tvaru.
Nyní si musíme vyčíslit i druhou matici $\mathbb{X}^T\mathbb{Y}$. Dostáváme

$$\mathbb{X}^T\mathbb{Y}=\begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}=\begin{pmatrix} \sum\limits_{i=1}^{n}y_i \\ \sum\limits_{i=1}^{n}x_iy_i \end{pmatrix}.$$

Nyní máme vše připraveno k dosazení do vzorce na výpočet hodnot regresních koeficientů, po dosazení dostáváme

$$\begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix}=\begin{pmatrix} \frac{\frac{1}{n}\sum\limits_{i=1}^{n}x_i^2}{\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2} & -\frac{\frac{1}{n}\sum\limits_{i=1}^{n}x_i}{\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2} \\ \frac{-\frac{1}{n}\sum\limits_{i=1}^{n}x_i}{\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2} & \frac{1}{\sum\limits_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2} \end{pmatrix}\begin{pmatrix} \sum\limits_{i=1}^{n}y_i \\ \sum\limits_{i=1}^{n}x_iy_i \end{pmatrix}=$$

$$
= \begin{pmatrix} \dfrac{\frac{1}{n}\sum\limits_{i=1}^{n} x_i^2 \sum\limits_{i=1}^{n} y_i}{\sum\limits_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum\limits_{i=1}^{n} x_i\right)^2} - \dfrac{\frac{1}{n}\sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} x_i y_i}{\sum\limits_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum\limits_{i=1}^{n} x_i\right)^2} \\ \dfrac{-\frac{1}{n}\sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i}{\sum\limits_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum\limits_{i=1}^{n} x_i\right)^2} + \dfrac{\sum\limits_{i=1}^{n} x_i y_i}{\sum\limits_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum\limits_{i=1}^{n} x_i\right)^2} \end{pmatrix} = \begin{pmatrix} \dfrac{\frac{1}{n}\sum\limits_{i=1}^{n} x_i^2 \sum\limits_{i=1}^{n} y_i - \frac{1}{n}\sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} x_i y_i}{\sum\limits_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum\limits_{i=1}^{n} x_i\right)^2} \\ \dfrac{\sum\limits_{i=1}^{n} x_i y_i - \frac{1}{n}\sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i}{\sum\limits_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum\limits_{i=1}^{n} x_i\right)^2} \end{pmatrix} .
$$

Vidíme, že při použití obou postupů nám vyšly naprosto stejné výsledky, což není nijak překvapivé.

Bonus: V bonusu bude naším úkolem provést regresní diagnostiku, kterou jsme v úlohách *b)* a *c)* nedělali.

Začneme úlohou *b)*. K tomu, abychom mohli posoudit, zda jsou všechny předpoklady lineárního regresního modelu v tomto případě splněny, si potřebujeme vykreslit jednotlivé grafy, které jsme detailně popisovali v textu seriálu. Na Obr. 2 můžeme vidět naměřená data a proloženou funkci. Vidíme, že proložená funkce na naměřená data dobře sedí (správně prokládá naměřená data a nejsou vidět žádné známky různého rozptylu měřených dat). Na Obr. 3 můžeme vidět hodnoty residuí vynesené v grafu oproti hodnotám nezávisle proměnné. V tomto grafu můžeme vidět rovnoměrně rozprostřená residua okolo osy x. Na Obr. 4 můžeme vidět vynesené hodnoty residuí oproti hodnotám proložené funkce. Opět není vidět žádná známka podezřelého chování residuí, která jsou rovnoměrně rozprostřené kolem osy x. Nakonec ještě musíme podle Obr. 5 zkontrolovat nezávislost měřených hodnot. Na tomto grafu opět nevidíme nic podezřelého, všechna residua jsou náhodně rozprostřeny kolem počátku souřadnic s žádnou tendencí shlukování v jednotlivých kvadrantech. Na žádných z těchto grafů jsme nenašli žádnou známku toho, že by byl porušen některý z předpokladů lineárního regresního modelu. Na závěr tedy můžeme říci, že použití lineární regrese bylo v tomto příkladě bezproblémové a obdržené výsledky můžeme považovat za přesné.

Nyní se budeme zabývat úlohou *c)*. Graf naměřených hodnot a proložené funkce můžeme vidět na Obr. 1. Z tohoto grafu můžeme vidět, že prokládaná funkce na naměřená data sedí poměrně dobře, ale je možné, že budeme mít problém s porušením předpokladu homoscedasticity (můžeme si všimnout, že naměřené hodnoty v pravé části grafu mají větší rozptyl než data na levé straně grafu). Na Obr. 6 můžeme vidět graf residuí oproti hodnotám nezávisle proměnné. Tento graf potvrzuje naše podezření o tom, že měřená data nemají stejný rozptyl. I zde je vidět, že residua v pravé části grafu mají větší rozptyl než residua v levé části grafu. Na Obr. 7 můžeme vidět graf residuí oproti hodnotám proložené funkce. I na tomto grafu můžeme pozorovat, že měřená data nemají stejný rozptyl. Na druhou stranu si ale musíme uvědomit, že rozdíl mezi rozptyly není nijak podstatně velký. Na Obr. 8 můžeme vidět graf residuí oproti posunutým residuím. Opět můžeme pozorovat, že residua nemají tendenci shlukovat se v žádném z kvadrantů a jsou rovnoměrně rozptýlena kolem počátku souřadnic, což značí, že naše měření byla nezávislá. Na závěr můžeme říci, že jsme z grafů sice vypozorovali, že naše meření neměla stejný rozptyl, ale porušení tohoto předpokladu není nijak výrazné a bude mít spíše zanedbatelný vliv na platnost výsledků obdržených použitím lineární regrese.

Na závěr ještě provedeme statistický test vhodnosti prokládané funkce v příkladu *c)* (v příkladu *b* jej provést nelze, neboť tam nemáme více měření pro stejné hodnoty nezávisle proměnné). Tento test byl popsán v textu seriálu a proto ho zde nebudeme podrobně opakovat, jen uvedeme, že testová hypotéza a alternativa mají tvar

$$H : \text{Prokládaná funkce je zvolena správně.}$$
$$A : \text{Prokládaná funkce není zvolena správně.}$$

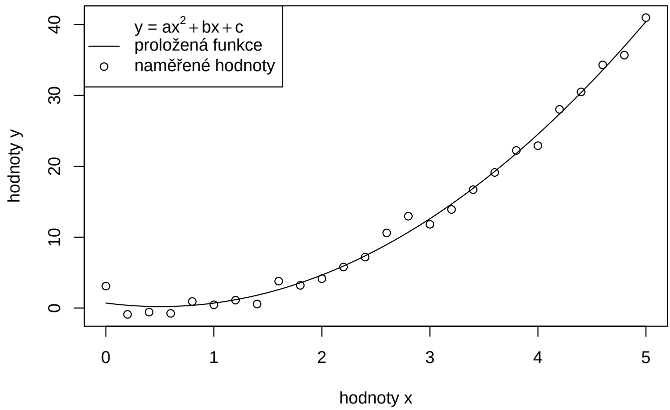
Použitím výpočetního softwaru R dostaneme hodnotu testové statistiky pro naše konkrétní data

$$CH \doteq 1{,}4$$

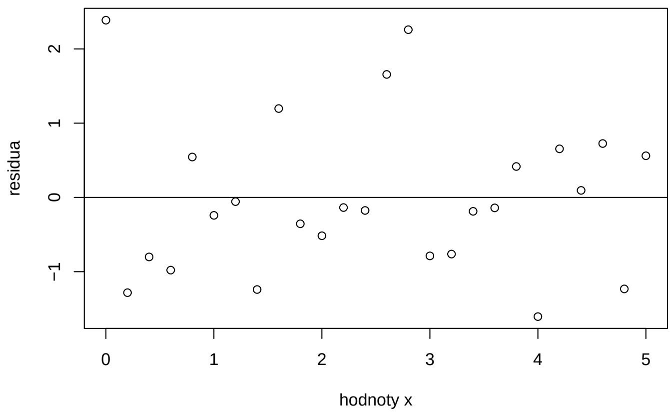
a p-hodnotu

$$p \doteq 0{,}24\,.$$

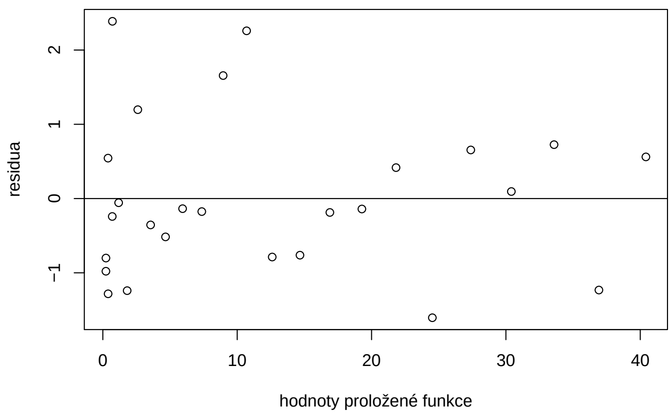
Vidíme tedy, že nezamítáme testovanou hypotézu na hladině spolehlivosti $\alpha = 0{,}05$. Toto značí (společně s grafy popsanými výše), že jsme prokládanou funkci zvolili správně.



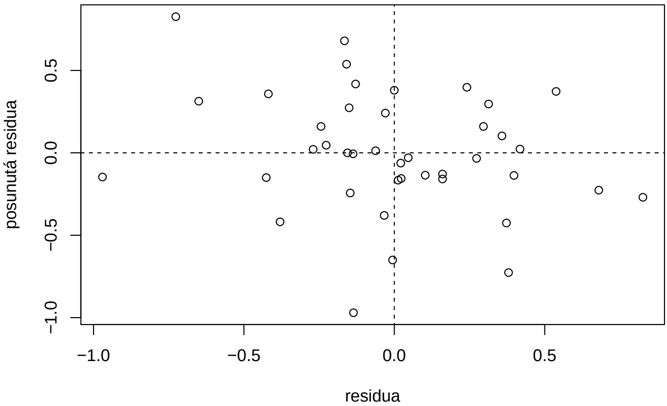
Obr. 2: Graf naměřených hodnot a proložené funkce z příkladu *b)*.



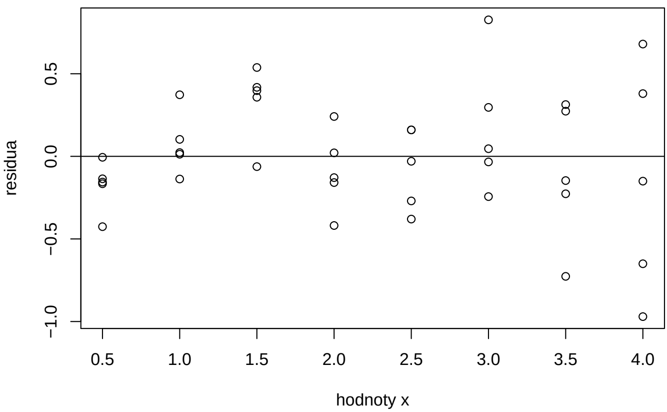
Obr. 3: Graf residuí oproti hodnotám nezávisle proměnné z příkladu *b)*.



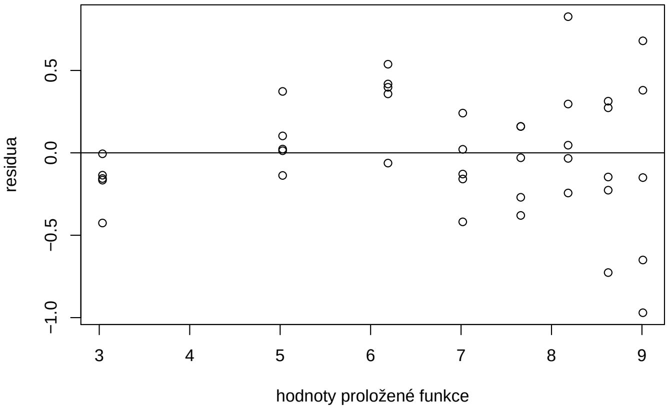
Obr. 4: Graf residuí oproti hodnotám proložené funkce z příkladu *b)*.



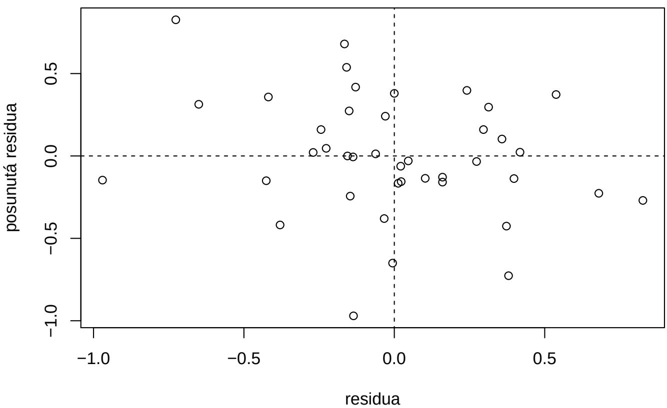
Obr. 5: Graf residuí oproti posunutým residuím z příkladu *b)*.



Obr. 6: Graf residuí oproti hodnotám nezávisle proměnné z příkladu *c)*.



Obr. 7: Graf residuí oproti hodnotám proložené funkce z příkladu *c)*.



Obr. 8: Graf residuí oproti posunutým residuím z příkladu *c)*.

Michal Nožička
nozicka@fykos.cz