

Úvodem

Milí FYKOSáci,

školní rok se přehoupl do posledního čtvrtletí a s ním přichází šestá a poslední série tohoto ročníku. Někteří z vás mají již nyní diplom úspěšného řešitele jistý a mnozí k němu mají slibně nakročeno. Pokud tedy chcete využít jedné z možností prominutí přijímacích zkoušek na Matfyz nebo pokud si prostě chcete dát diplom za rámeček do výstavy, máte poslední šanci zvýšit svůj bodový zisk. Venku všechno kvete a voní a my přicházíme s letos posledním zadáním, stejně jako s řešenými úlohami z minulé série. Za některá chybějící řešení se hluboce omlouváme. Spoustu výtečných nápadů při řešení přejí vaši

Organizátoři



Zadání VI. série



Termín uploadu: 16. 5. 2017 23.59

Termín odeslání: 15. 5. 2017

Úloha VI.1 ... dost těžké kulometry

3 body

Na auto připevníme dopředu dva kulometry, které vystřelují kulky o hmotnosti $m = 25\text{ g}$ rychlostí $v_1 = 500\text{ m}\cdot\text{s}^{-1}$, každý s frekvencí 10 výstřelů za sekundu. Auto se rozjede po rovině rychlostí $v_2 = 80\text{ km}\cdot\text{h}^{-1}$ a poté začne střílet. Kolik nábojů vystřelíme, než auto zastaví? Během palby nepřidáváme plyn, odpor vzduchu a kol zanedbáváme. Tepelné ztráty uvnitř zbraní jsou taktéž zanedbatelné.

Úloha VI.2 ... upadlo

3 body

Z jaké výšky nad povrchem neutronové hvězdy bychom museli „upustit“ předmět, aby dopadl na její povrch v rychlosti $0,1 c$ ($0,1$ rychlosti světla). Naše neutronová hvězda má hmotnost $1,5$ násobku hmotnosti Slunce a průměr $d = 10\text{ km}$. Zanedbejte atmosféru neutronové hvězdy a její rotaci. Zanedbejte relativistické korekce. Srovnajte ale, jakého výsledku byste dosáhli, pokud by pád probíhal v homogenním gravitačním poli (které má intenzitu stejnou jako na povrchu planety) s tím, kdy pád probíhá v radiálním gravitačním poli.

Bonus: Uvažujte korekci na speciální teorii relativity v případě pádu v homogenním poli.

Úloha VI.3 ... relativistický Zenonův paradox

6 bodů

Superman a Flash se rozhodli, že si dají závod. Závod se koná v hlubokém vesmíru, protože na Zemi není dostatečně dlouhá rovná pláž. Flash, protože je pomalejší, startuje s délkovým náskokem l před Supermanem. Flash v jednu chvíli vyběhne s konstantní rychlostí v_F srovnatelnou s rychlostí světla. Ve chvíli, kdy si Superman všimne, že Flash vyběhl, vyběhne také, a to konstantní rychlostí $v_S > v_F$. Za jak dlouho Superman Flashe dožene (z pohledu Supermana)? A za jak dlouho Flashe dožene Superman (z pohledu Flashe)? A byl vůbec závod spravedli-

vě odstartován, resp. dokázali byste vymyslet spravedlivější způsob (příčemž náskok l má být ponechán)?

Úloha VI.4 ... zastřel si svého potkana

7 bodů

Mírek by rád zastřelil potkana, kterého vídá na kolejích. Připravil si tedy jednoduchou vzduchovou pušku, kterou si můžeme modelovat jako trubku s konstantním průřezem $S = 15 \text{ mm}^2$ a délkou $l = 30 \text{ cm}$, která je na jedné straně uzavřená a na druhé otevřená. Do ní se chystá Mírek umístit náboj hmotnosti $m = 2 \text{ g}$, který trubku akorát utěsní, a to ve vzdálenosti $d = 3 \text{ cm}$ od uzavřeného konce. Náboj zde zatím nechá upevněný v klidu a natlakuje uzavřenou část trubky na určitý tlak p_0 . Posléze náboj uvolní. Chce, aby na konci ústí byla rychlost náboje minimálně $v = 90 \text{ m}\cdot\text{s}^{-1}$. Poradte mu, na jaký tlak by musel vzduchovou pušku natlakovat, aby náboj vyšel s takovou rychlostí, pokud by plyn byl ideální, a diskutujte realističnost uspořádání. Předpokládejte, že náboj je uvolňován kvazistatickým adiabatickým dějem, kde $\kappa = 7/5$, protože se jedná o dvouatomový plyn. Uvažujte, že z vnějšku působí na náboj atmosférický tlak $p_a = 10^5 \text{ Pa}$. Zanedbejte energetické ztráty vyvolané třením, odporem vzduchu a stlačováním plynu před nábojem.

Úloha VI.5 ... přetáhni ho přes prsty

8 bodů

Máme homogenní tyč konstantního průřezu délky l připevněnou na jednom konci k otočnému kloubu. Na počátku směřuje tyč přímo vzhůru a jsme v homogenním tíhovém poli velikosti g . Tyč se vlivem mírného závanu větru začne otáčet a „padat“ dolů, ale stále je držena otočným kloubem. S jakým zrychlením se bude pohybovat konec tyče v průběhu času?

Úloha VI.P ... vypařující se asteroid

9 bodů

Umístíme hodně velký kus ledu, dejme tomu o průměru 1 km , do blízkosti hvězdy podobné Slunci na kruhovou dráhu. Blízkost je tak velká, že rovnovážná teplota černého tělesa by v této vzdálenosti byla zhruba $30 \text{ }^\circ\text{C}$. Co se bude dít s takovým asteroidem a jeho dráhou? Asteroid nemá vázanou rotaci.

Úloha VI.E ... skladba jako od Cimrmana

12 bodů

Sežeňte si skleničku na víno, ideálně tenkou se zabroušeným okrajem. Nejprve změřte vnitřní průměr skleničky v závislosti na výšce ode dna. Pak ji rozeznívejte, ideálně navlhčeným prstem pohybem po jejím okraji – někdy to chce trochu trpělivosti. Změřte závislost frekvence tónů, které sklenička vydává v závislosti na výšce naplnění vody v ní (alespoň pro 5 hladin vody a dvě frekvence v každé výšce).

Nápověda: Pokud je sklenička tenkostěnná, můžete její vnitřní rozměry považovat za stejné jako vnější a díky tomu závislost jejího průměru na výšce určit z vhodné fotografie s měřítkem. Pro měření zvuku doporučujeme freeware program Audacity (Rozbor → Kreslit spektrum).

Úloha VI.S ... nelineární

10 bodů

a) Zkuste vlastními slovy popsat, k čemu a jak se používá nelineární regrese (postačí vlastními slovy popsat následující: model nelineární regrese, způsob odhadu regresních koeficientů,

vyjádření nejistot odhadů regresních koeficientů a hodnot prokládané funkce, statistické testy hodnot regresních koeficientů, identifikovatelnost parametrů a způsob volby prokládané funkce). Není potřeba uvádět přesná matematická odvození, stačí požadované pojmy a vlastnosti stručně popsat.

- b) V příloženém datovém souboru *regrese1.csv* naleznete dvojice hodnot (x_i, y_i) . Těmito daty chceme proložit teoretickou funkční závislost, kterou je v tomto případě sinusoida, tedy funkce tvaru

$$f(x) = a + b \cdot \sin(cx + d).$$

Vykreslete graf naměřených hodnot a proložené funkce a stručně ho okomentujte (takovýto graf musí mít všechny náležitosti). Není potřeba dělat regresní diagnostiku.

Nápověda: Dejte si pozor na identifikovatelnost parametrů v tomto modelu a vhodné omezující podmínky na parametr c .

- c) V příloženém datovém souboru *regrese2.csv* naleznete dvojice hodnot (x_i, y_i) . Těmito daty chceme proložit teoretickou funkční závislost, kterou je v tomto případě exponenciála, tedy funkce tvaru

$$f(x) = a + e^{bx+c}.$$

Určete hodnoty odhadů všech regresních koeficientů včetně nejistot měření.

Nápověda: Grafickou metodou ověřte předpoklad homoskedasticity a v případě potřeby pro určení nejistot měření regresních koeficientů použijte Whiteův (sendvičový) odhad kovarianční matice.

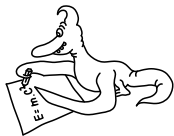
- d) V příloženém datovém souboru *regrese3.csv* naleznete dvojice hodnot (x_i, y_i) . Těmito daty chceme proložit teoretickou funkční závislost, kterou je v tomto případě hyperbola, tedy funkce tvaru

$$f(x) = a + \frac{1}{bx + c}.$$

Vykreslete graf naměřených dat v podobě průměrů a chybových úseček a proložené funkce a stručně ho okomentujte (takovýto graf musí mít všechny náležitosti). Proveďte regresní diagnostiku.

Bonus: V příloženém datovém souboru *regrese4.csv* naleznete dvojice hodnot (x_i, y_i) . Těmito daty chceme proložit teoretickou závislost, která je ovšem příliš složitá na analytické vyjádření. Proložte těmito daty regresní spliny (s vhodně zvolenými uzly a vhodně zvoleným stupněm).

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



Řešení V. série

Úloha V.1 ... vesmírný sněhulák

3 body; průměr 3,17; řešilo 52 studentů

Jakou silou bude přidržována hlava našeho sněhuláka, který si volně poletuje ve vesmíru? Máme sněhuláka tvořeného pouze homogenními koulemi o hustotě ρ , jejichž středy leží na jedné přímce a koule se dotýkají, jsou umístěné v pořadí od největší po nejmenší a s tím, že nejmenší koule (hlava) má poloměr r a každá další má dvojnásobný poloměr, co ta předchází. Ve vesmíru je pouze náš sněhulák a nijak nerotuje.

Bonus: Zobecněte úlohu pro počet koulí $N \geq 3$. Bude se síla blížit nějaké konečné hodnotě pro $n \rightarrow \infty$, nebo půjde k nekonečnu?

Karel vymýšlel úlohu na Fyziklání a pak si řekl, že by ten výsledek nechtěl kontrolovat.

Řešme rovnou bonusovou úlohu a nakonec dostaneme výsledek základní verze dosazením $N = 3$. Koule si označme čísly od nejmenší 1 po největší N . Poloměr i -té koule bude $r_i = 2^{i-1}r$. Hmotnost i -té koule bude

$$m_i = \rho V_i = \frac{4}{3}\pi\rho r_i^3 = \frac{4}{3}\pi\rho 2^{3(i-1)}r^3.$$

Síla, kterou hledáme, je součtem jednotlivých sil

$$F_{\text{celk}} = \sum_{i=2}^N F_{1i} = \sum_{i=2}^N G \frac{m_1 m_i}{d_{1i}^2},$$

kde F_{1i} je gravitační síla mezi první a i -té kouli (kde $i \geq 2$) a d_{1i} je vzdálenost jejich středů. Vzdálenost středů už není úplně triviální. Z první a i -té koule máme jeden poloměr a z koulí mezi nimi dvojnásobek poloměru. Rozepsat si to můžeme jako

$$d_{1i} = r_1 + 2(r_2 + \dots + r_{i-1}) + r_i = r(1 + 2(2 + 4 + \dots + 2^{i-2}) + 2^{i-1}).$$

Nyní využijeme vzorec pro součet konečné geometrické řady

$$\sum_{k=0}^K aq^k = a \frac{q^{k+1} - 1}{q - 1}.$$

$$d_{1i} = r \left(1 + 4 \frac{2^{i-2} - 1}{2 - 1} + 2^{i-1} \right) = r(3 \cdot 2^{i-1} - 3) = 3r(2^{i-1} - 1).$$

Nyní můžeme dosadit všechno do síly

$$\begin{aligned} F_{\text{celk}} &= \sum_{i=2}^N G \frac{\frac{4}{3}\pi\rho r^3 \frac{4}{3}\pi\rho r_i^3}{(3r(2^{i-1} - 1))^2} = G \sum_{i=2}^N \frac{2^4 \pi^2 \rho^2 r^4}{3^4} \frac{2^{3(i-1)}}{(2^{i-1} - 1)^2} = \\ &= \frac{\pi^2 \rho^2 G r^4}{3^4} \sum_{i=2}^N \frac{2^{3i+1}}{(2^{i-1} - 1)^2} \end{aligned}$$

Nejdříve dořešíme bonusové otázky. Jednak prakticky ihned vidíme, že pro $i \rightarrow \infty$ jde $F_{1i} \rightarrow \infty$, dokonce roste řádově jako 2^i . Tím pádem součet těchto členů půjde také do nekonečna. Pokud se pokusíme udělat nějaký částečný součet pro obecné N , tak když to například necháme spočítat Wolfram Mathematicu, tak dostaneme sumu nějakých ošklivých polynomů generovaných digamma funkcí a jejími derivacemi, takže nic pěkného. Takže nám bude stačit ještě ten relativně pěkný zápis, u kterého jsme skončili.

Řešením základní úlohy je

$$F_{123} = \frac{\pi^2 \rho^2 Gr^4}{3^4} \sum_{i=2}^3 \frac{2^{3i+1}}{(2^{i-1} - 1)^2} = \frac{\pi^2 \rho^2 Gr^4}{3^4} \left(128 + \frac{1024}{9} \right) = \frac{2^7 \cdot 17}{3^6} \pi^2 \rho^2 Gr^4 \\ \doteq \rho^2 r^4 \cdot 2,0 \cdot 10^{-9} \text{ N} \cdot \text{kg}^{-2} \cdot \text{m}^2.$$

Hustotu ani poloměr jsme neměli zadanou. Nicméně můžeme si říci, že máme třeba sněhuláka železného $\rho \doteq 7900 \text{ kg} \cdot \text{m}^{-3}$ a nejmenší koule má poloměr $r = 1 \text{ m}$. Tedy na Zemi oproti člověku by to byl docela velký sněhulák, ale zase ve srovnání s velikostí planety by byl malý. V tom případě by hlava držela silou $F_{123} \doteq 0,12 \text{ N}$. Tedy držela by, ale slabě. Dokud nemáme v našem vesmíru nic jiného, ani průvan, ani makroskopickou elektromagnetickou interakci, tak by ta hlava držela. Samozřejmě kdyby někdo do hlavy trochu strčil, tak by pak sjela a nebyl by to už sněhulák, ale tři koule navzájem se dotýkající.

Karel Kolář
karel@fykos.cz

Úloha V.2 ... koule ve vazkých tekutinách 3 body; průměr 2,98; řešilo 43 studentů

V některých případech řešení úloh s odporem vzduchu či obecně tekutiny používáme pro odporovou sílu Newtonův vzorec $F = C \rho S v^2 / 2$, kde C je součinitel odporu tělesa ve směru pohybu tělesa, ρ je hustota tekutiny, S je průřez a v je rychlost pohybu tělesa. Ten obvykle docela dobře platí pro turbulentní prostředí. Zajímáme se o kouli, pro kterou $C = 0,50$. V laminárním proudění pak obvykle používáme Stokesův vztah $F = 6\pi\eta r v$, kde η je dynamická viskozita tekutiny a r je poloměr koule. Pokud máme nějakou konkrétní kouli, je možné, aby se pro nějakou rychlost tyto odpory rovnaly? Jak bude tato rychlost záviset na poloměru koule?

Karel na konferenci zaslechl, že lidi mají problémy s rovnostmi.

Vzhledem k tomu, že otázkou je, jestli se mohou dva výrazy rovnat, tak matematickou odpověď na takovou otázku je obvykle ano, pokud nedojde k nějakému sporu. Proto dáme síly do rovnosti s uvážením, že průřez koule je $S = \pi r^2$

$$\frac{1}{2} C \rho S v^2 = 6\pi\eta r v \quad \Rightarrow \quad v_0 = 0 \quad \vee \quad v = \frac{12\pi\eta r}{C \rho S} = \frac{24\eta}{\rho r}.$$

Zdá se, že jsme k nějakému sporu nedošli. Vypadá to, že požadovaná rychlost bude buď nulová¹, nebo nepřímo úměrná poloměru koule. Triviální řešení, tedy že při nulové rychlosti budou obě síly nulové, je zřejmě správné. Pro jistotu si ještě můžeme dosadit hodnoty v netriviálním řešení, pro nějakou takovou běžnou kouli a běžnou tekutinu. Vybereme si železnou

¹Zjevně pak na velikosti koule nebude záležet.

kuličku s poloměrem $r = 1,0 \text{ cm}$, která je ponořená do vody o hustotě $\rho = 1,00 \text{ g}\cdot\text{cm}^{-3}$. Voda má dynamickou viskozitu $\eta = 1,0 \cdot 10^{-3} \text{ Pa}\cdot\text{s}$ při pokojové teplotě. Výsledná rychlost je pak $2,4 \cdot 10^{-3} \text{ m}\cdot\text{s}^{-1}$. To je relativně nízká rychlost. Pro vyšší rychlosti bude odhad odporové síly vyšší pro Newtonův vztah a pro nižší rychlosti bude vyšší pro Stokesův vztah.

Ve skutečnosti bychom neměli prakticky používat ve stejné oblasti oba vzorce, protože jeden je vhodný popis pro laminární proudění a druhý pro turbulentní. Ostatně vidíme, že bychom měli používat zpravidla ten odpor, který má vyšší hodnoty. Turbulentní se totiž stává proudění pro vysoké rychlosti a laminární je naopak v nízkých rychlostech. I když toto tvrzení není také úplně přesné, protože rychlost, která nám vyšla pro rovnost u koule, odpovídá řádově desítkám v hodnotě Reynoldsova čísla. To odpovídá ještě laminárnímu proudění a za turbulentní se považuje obvykle proudění od cca od hodnoty 1000 Reynoldsova čísla či o něco vyšší.

Komentáře k došlým řešením

Hodnocení bylo mírné, i vzhledem k tomu, že jsme v původním vzorovém řešení také zapomněli uvést jako jedno z řešení nulovou rychlost. Kdo na to nezapomněl a k tomu svoje řešení dobře okomentoval, mohl snadno získat bonusový bod.

Karel Kolář
karel@fykos.cz

Úloha V.P ... sklíčka

8 bodů; průměr 4,90; řešilo 29 studentů

Popište zobrazovací soustavy mikroskop (složený ze 2 spojek) a Keplerův dalekohled. Vysvětlete rozdíl ve funkci a konstrukci mikroskopu a dalekohledu a načrtněte průchod paprsků. Jak se dá smysluplně definovat zvětšení pro dané optické prvky? Odvoďte pro zvětšení konkrétní vzorce.

Kuba konečně pochopil, jak to všechno funguje!

Nejprve se podívejme, co musí obraz předmětu splňovat, abychom ho mohli vidět. V oku je spojná čočka, která z rozbíhajících se paprsků vytvoří skutečný obraz na sítnici. Současně je pro oko nejpohodlnější, když je obraz v dáli, protože nemusí být akomodováno. Proto, abychom něco v libovolné zobrazovací soustavě viděli, měl by být obraz ideálně virtuální a v nekonečnu. Jiný obraz (virtuální nebo reálný) můžeme také pozorovat, ale oko musí zaostřit tak, aby ho zobrazilo přesně na sítnici a nikam jinam.

Základní princip soustavy dvou spojek (Keplerův dalekohled se také skládá ze dvou spojek) tedy bude, že první čočka (objektiv) zobrazí předmět do ohniskové roviny druhé čočky (okuláru) a ta poté vytvoří obraz v nekonečnu². Hlavní rozdíl je, že mikroskopem se zvětšují blízké předměty (tedy se jedná o zobrazení z blíзка do nekonečna) a dalekohledem předměty vzdálené (zobrazení z nekonečna do nekonečna). Objektiv dalekohledu tedy vytvoří meziobraz ve své ohniskové rovině, kam musíme umístit také ohnisko okuláru.

Oproti tomu objektiv mikroskopu vytvoří meziobraz dále než ve své ohniskové rovině, a to podle zobrazovací rovnice

$$\frac{1}{a_1} + \frac{1}{a'_1} = \frac{1}{f_1},$$

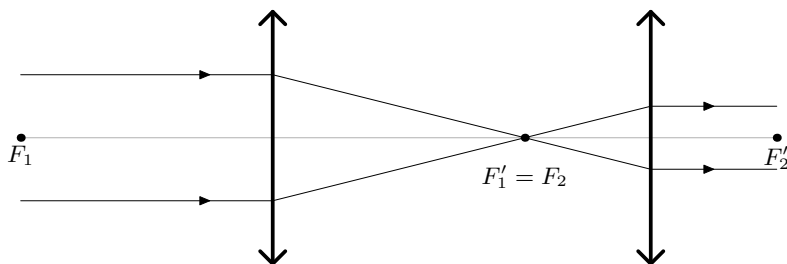
² Takový obraz není virtuální v pravém slova smyslu, ale tento případ lze interpretovat jednak tak, že se vytvoří skutečný obraz v plus nekonečnu, nebo že se vytvoří virtuální obraz v minus nekonečnu, což je přesně to, co chceme.

kde a_1 značí vzdálenost předmětu od objektivu, a'_1 vzdálenost meziobrazu od objektivu a f_1 ohniskovou vzdálenost objektivu. Mikroskop má typicky fixní vzdálenost objektivu a okuláru, a co se dá měnit, je vzdálenost a_1 . Označme Δ vzdálenost bližších ohnisek objektivu a okuláru tak, že je kladná, pokud je ohnisko F'_1 nalevo od ohniska F_2 (pro dalekohled by platilo $\Delta = 0$). Potom můžeme zobrazovací rovnici psát ve tvaru

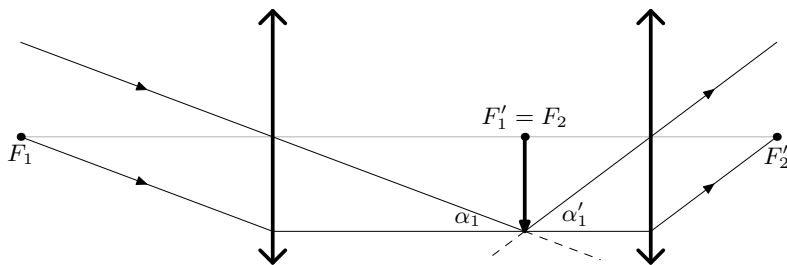
$$\frac{1}{a_1} + \frac{1}{f_1 + \Delta} = \frac{1}{f_1}, \quad (1)$$

odkud lze vypočítat patřičnou vzdálenost a_1 .

Kromě hvězdářských dalekohledů jsou ještě dalekohledy na pozemské vzdálenosti, které se však od obou soustav ze zadání liší – zobrazujeme předměty, ke kterým se můžeme dostat relativně blízko, a ohniska okuláru a objektivu nesplývají. Je tady jistá podobnost s mikroskopy. V tomto případě však nelze měnit vzdálenost a_1 , proto se mění vzdálenost Δ , kterou můžeme vypočítat se stejné rovnice (1).



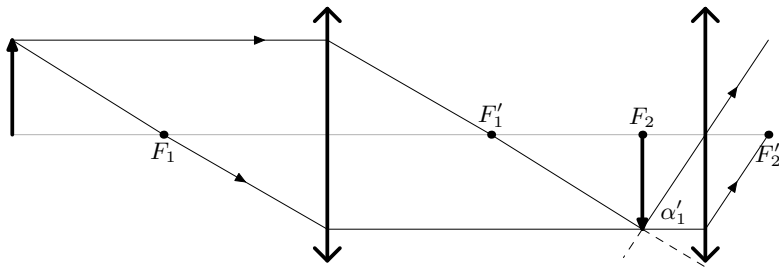
Obr. 1: Průchod horizontálních paprsků dalekohledem



Obr. 2: Průchod šikmých paprsků dalekohledem

Na obr. 1 a 2 vidíme průchod paprsků dalekohledem a na obr. 3 průchod mikroskopem. Současně je na obrázcích vidět, že do dalekohledu přichází paprsky z nekonečna a ohniska objektivu a okuláru splývají. Dále je předmět blízko k mikroskopu a na obr. 3 je vyznačena vzdálenost Δ .

Smysluplné zvětšení je poměr úhlových velikostí, a to z jednoduchého důvodu. Velikost obrazu vytvořeného na sítnici oka závisí právě na úhlové velikosti předmětu. Pro mikroskop (lupu a další zobrazovací soustavy na blízko) se porovnává úhel obrazu v mikroskopu a úhel



Obr. 3: Průchod paprsků mikroskopem

předmětu v konvenční zrakové vzdálenosti $L = 25$ cm. Tato vzdálenost je zvolena proto, že čím blíže předmět je, tím větší ho vidíme (a zvětšení chceme porovnávat s nejlepším možným pozorováním pouhým okem), ale současně oko není schopné zaostřit libovolně blízké předměty. Konvenční zraková vzdálenost je blízkým bodem (nejblíže bod, na který jsme schopni zaostřit) průměrného lidského jedince.

V případě dalekohledu nemůžeme pozorovat předměty v konvenční zrakové vzdálenosti, proto je třeba zvětšení definovat jako poměr úhlových velikostí tělesa v dalekohledu a bez dalekohledu. To můžeme udělat díky tomu, že vzdálenost pozorovaných předmětů je výrazně vyšší, než jak moc jsme schopni se k těm předmětům přiblížit (tedy toto zvětšení nezávisí na naší poloze).

Při výpočtu zvětšení předpokládáme, že úhlové velikosti předmětu i obrazu jsou malé, tedy můžeme použít vzorec

$$\begin{aligned}\sin x &\approx x, \\ \operatorname{tg} x &\approx x.\end{aligned}$$

Tuto aproximaci můžeme použít, protože mikroskopem koukáme na velice malé předměty a dalekohledem koukáme do velmi velké vzdálenosti. Jedná se o klasickou paraxiální aproximaci.

Spočtíme nyní zvětšení mikroskopu. Úhlová velikost předmětu je

$$\alpha = \frac{y}{L},$$

kde y je jeho příčný rozměr. Dále z podobnosti trojúhelníků na obr. 3 vidíme

$$\frac{y}{f_1} = \frac{-h}{\Delta} = \frac{-f_2 \alpha'}{\Delta},$$

kde h je příčná velikost meziobrazu (který je převrácený). Teď už můžeme psát vzorec pro úhlové zvětšení mikroskopu

$$\Gamma = \frac{\alpha'}{\alpha} = \frac{-\frac{y\Delta}{f_1 f_2}}{\frac{y}{L}} = -\frac{L\Delta}{f_1 f_2}.$$

K výpočtu zvětšení dalekohledu využijeme obr. 2. Interpretovat ho můžeme dvojím způsobem, buď je předmět vycentrovaný na optické ose a paprsky přichází od jeho horního kraje (a y je tedy jeho poloměr), nebo dolní okraj předmětu je na optické ose, paprsky také přichází

od jeho horního kraje, ale y je jeho celý průměr. Úhel paprsků a optické osy odpovídá úhlu α a úhel finálních paprsků, mířících do oka, s optickou osou je právě úhel α' . Z obrazku je vidět

$$h = f_1\alpha = f_2\alpha',$$

kde jsme však nechali pro názornost pouze kladná znaménka. Protože bude obraz převrácený, musíme do vzorce pro zvětšení přidat minus. Nyní jednoduše vyjádříme zvětšení dalekohledu

$$\Gamma = -\frac{\alpha'}{\alpha} = -\frac{f_1}{f_2}.$$

Aby tedy dalekohled skutečně zvětšoval, musí mít okulár menší ohniskovou délku než objektiv. Nenechte se však zmást tím, že v tomto případě bude obraz předmětu dalekohledem příčně zmenšený (jak vidíme na obr. 1), záleží totiž především na úhlovém zvětšení.

Reálně využívané zobrazovací soustavy jsou složeny z mnoha zrcadel, hranolů a čoček, které převrací obraz a především eliminují vady zobrazení – např. sférickou a chromatickou.

Komentáře k došlým řešením

Mnoho z vás řešilo úlohu jako běžnou číslovanou. Hlavní pointou zde bylo např. odůvodnit, že příčné zvětšení není určující, že ve skutečnosti nás zajímá úhlové, nebo v rámci lepšího pochopení zobrazovací optiky vztahy pro zvětšení odvodit (finální vzorec se běžně učí na středních školách). Také jste mohli vymyslet, že zaostřování mikroskopu spočívá v tom, že meziobraz musí vzniknout právě v ohnisku okuláru (nikoli např. v umístění předmětu do ohniska objektivu). Další zajímavý problém je, kde se vzala konvenční zraková vzdálenost, a proč je pro zvětšení použita právě vzdálenost blízkého bodu lidského oka.

Stejně tak mnoho z vás překreslilo (nebo zkopírovalo) schéma mikroskopu z wikipedie, aniž by ho řádně ocitovala. Ve skutečnosti si na něm ale můžete všimnout chyby, že pokud meziobraz vznikne v ohnisku okuláru, musí podle zobrazovací rovnice vzniknout obraz v nekonečnu, a tedy paprsky vycházející z mikroskopu musí být rovnoběžné. Lépe řečeno, toto je teoretická idea a prakticky se nám to nepovede nastavit přesně.

Jakub Dolejší
dolejsi@fykos.cz

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 funkcí, 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 rádně 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 $(x_1, y_1), \dots, (x_n, y_n)$ 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 nepřímou 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

$$u_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 f_1(x_i) - \dots - \hat{\beta}_k f_k(x_i),$$

kde $\hat{\beta}_0, \dots, \hat{\beta}_k$ 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. To se může následně ověřovat na vhodných grafech residuí (např. residua oproti nezávisle proměnné, residua oproti hodnotám proložené funkce, residua oproti posunutým residuům atd.).

- b) Ve výpočetním prostředí 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 uveďte je na výstupu. Pro tato konkrétní data a prokládanou funkci dostaneme odhad regresního koeficientu u x^2

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

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

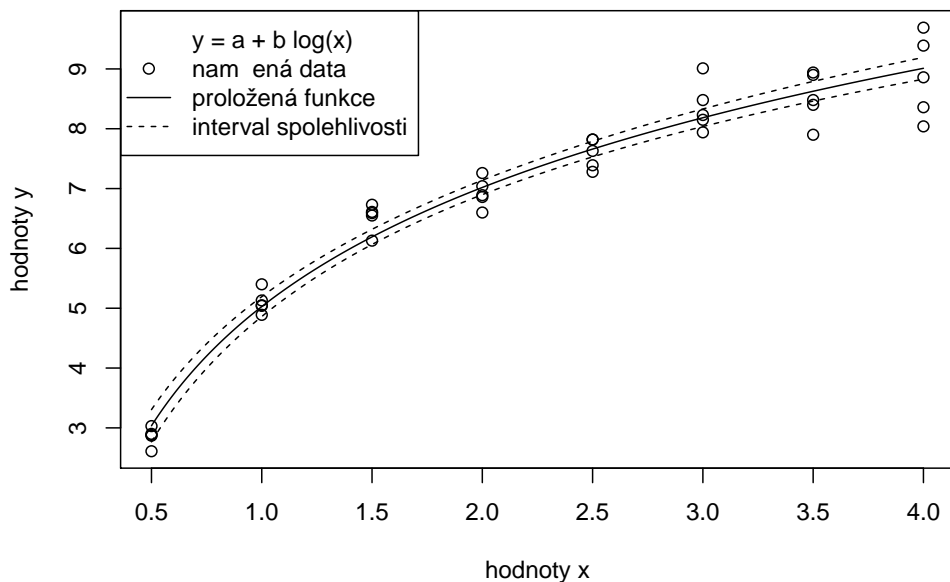
$$f(x) = a + b \cdot \log(x).$$

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

$$a = (5,0 \pm 0,08),$$

$$b = (2,87 \pm 0,09),$$

kde využíváme klasické zkrácený zápis nejistot měření. Na Obr. 4 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.



Obr. 4: Graf s naměřenými daty, proloženou funkční závislostí a 95% intervalem spolehlivosti pro prokládanou funkci.

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 nezávisle 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

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

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

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

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

$$\begin{aligned} \sum_{i=1}^n y_i - na - b \sum_{i=1}^n x_i &= 0, \\ \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i - b \sum_{i=1}^n x_i^2 &= 0. \end{aligned}$$

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

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

Zpětným dosazením potom získáme vzorec pro odhad parametru a ve tvaru

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

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

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

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 $(\mathbb{X}^T \mathbb{X})^{-1}$. 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.

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

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

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

V průběhu těchto úprav jsme používali jen odečítání násobku jednoho řádku od druhého a násobení řádků konstantou. Konkrétně jsme v první úpravě odečetli $\frac{1}{n} \sum_{i=1}^n x_i$ násobek prvního řádku od druhého. Ve druhé úpravě jsme druhý řádek vydělili $\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2$, následně jsme odečetli $\sum_{i=1}^n x_i$ 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

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

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

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

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

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. 5 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. 6 můžeme vidět hodnoty residuí vynesené v grafu oproti hodnotám nezávisle proměnné. V tomto grafu můžeme vidět rovnoměrně rozprostřená residua okolo osy x . Na Obr. 7 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. 8 zkontrolovat nezávislost měřených hodnot. Na tomto grafu opět nevidíme nic podezřelého, všechna residua jsou náhodně rozprostřena 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 případě bezproblémové a obdržené výsledky můžeme považovat za přesné.

Nyní se budeme zabývat úlohou *c)*. Graf naměřených hodnot a proložené funkce můžeme vidět na Obr. 4. 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. 9 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. 10 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. 11 můžeme vidět graf residuí oproti posunutým residuům. 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 vypožorovali, že naše měření 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 : Prokládaná funkce je zvolena správně.

A : Prokládaná funkce není zvolena správně.

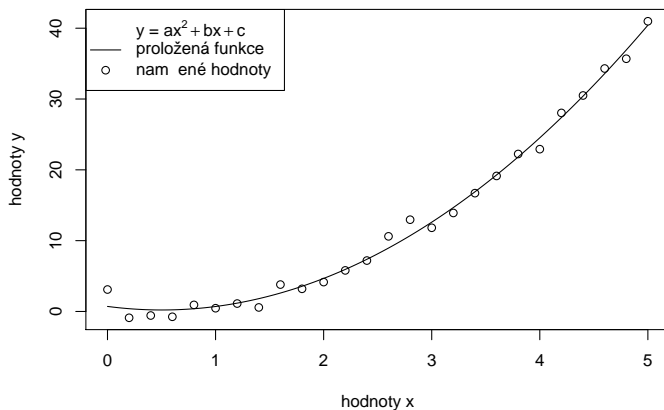
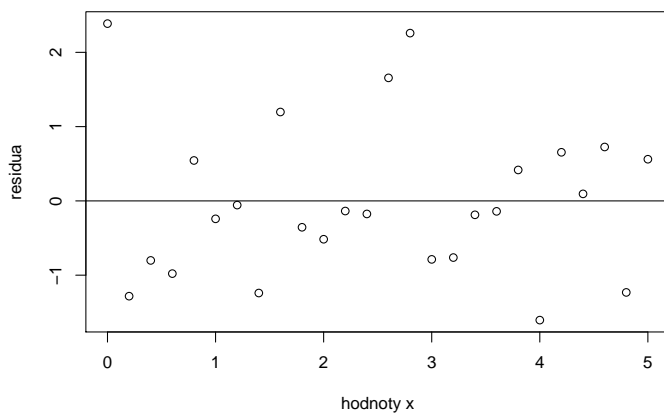
Použitím výpočetního softwaru R dostaneme hodnotu testové statistiky pro naše konkrétní data

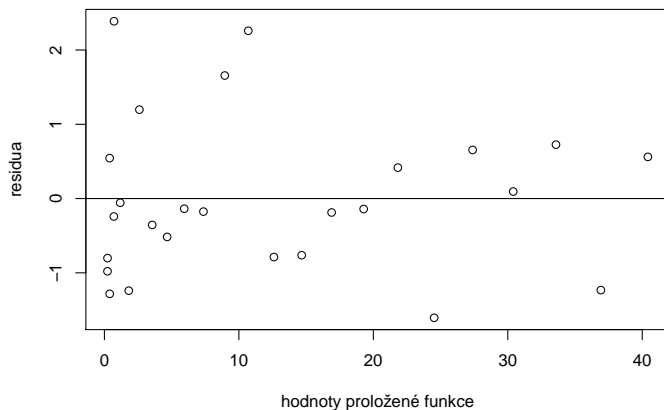
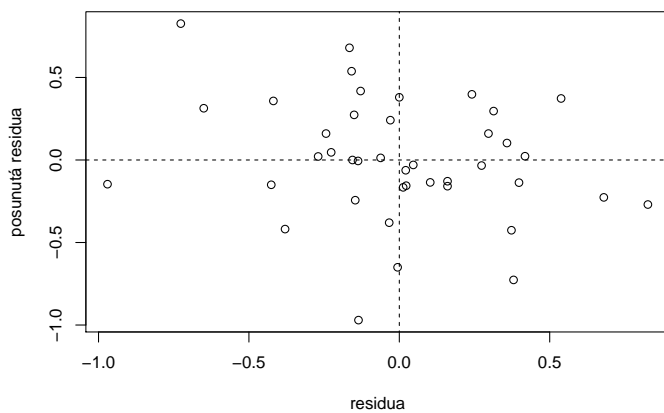
$$CH \doteq 1,4$$

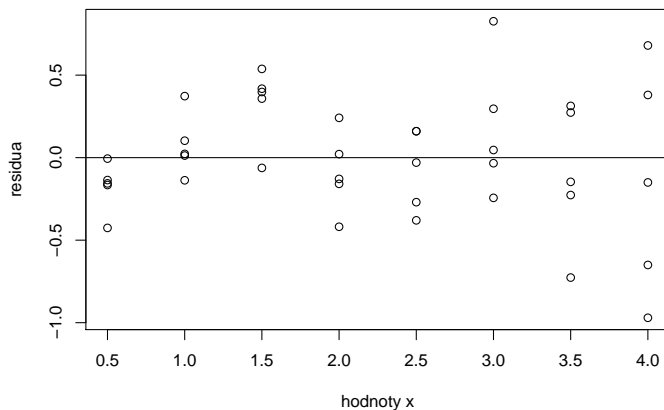
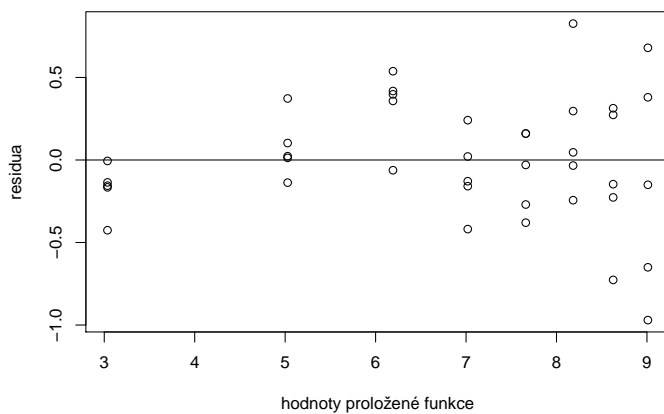
a p -hodnotu

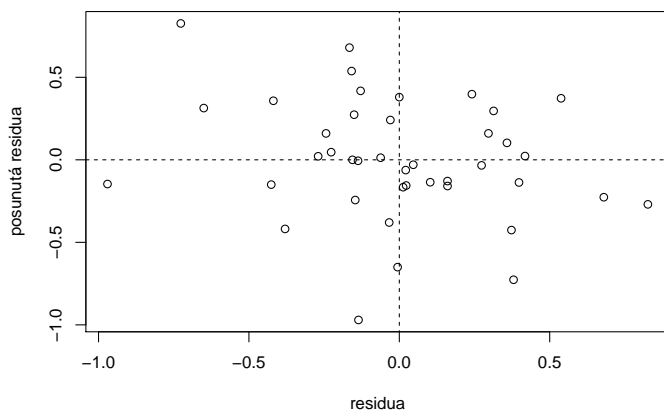
$$p \doteq 0,24.$$

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

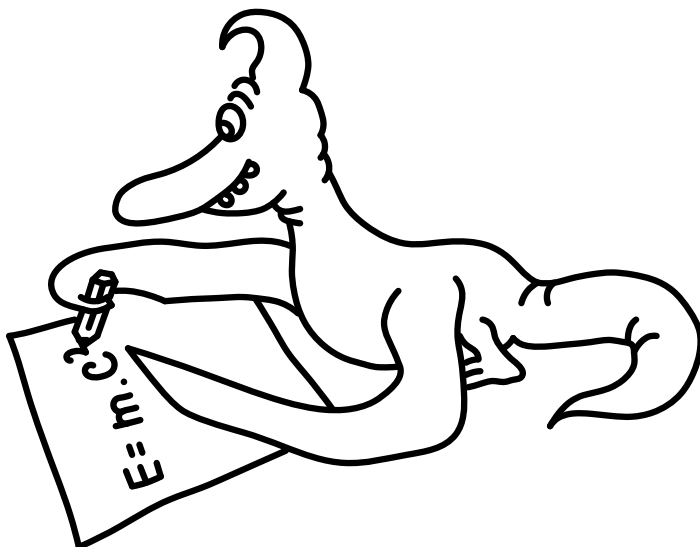
Obr. 5: Graf naměřených hodnot a proložené funkce z příkladu *b*).Obr. 6: Graf residuí oproti hodnotám nezávisle proměnné z příkladu *b*).

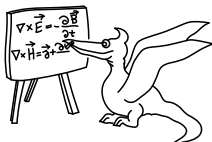
Obr. 7: Graf residuí oproti hodnotám proložené funkce z příkladu *b*).Obr. 8: Graf residuí oproti posunutým residuům z příkladu *b*).

Obr. 9: Graf residuí oproti hodnotám nezávisle proměnné z příkladu *c*).Obr. 10: Graf residuí oproti hodnotám proložené funkce z příkladu *c*).



Obr. 11: Graf residuí oproti posunutým residuům z příkladu *c*).





Seriál: Seriál - Zpracování dat 6. díl

V tomto posledním díle seriálu se budeme věnovat pokročilejším partiím regresní analýzy. Nejprve si popíšeme, jak vyřešit případy, kdy není zřejmé, jakou máme zvolit prokládanou funkci, což vyřešíme lokální aproximací polynomy. Dále si popíšeme případy, ve kterých potřebujeme naměřenými daty proložit funkční závislost, která není lineární v neznámých regresních parametrech. Poznáme, že z hlediska zpracování dat a práce s matematickým softwarem je nelineární regrese prakticky stejná jako regrese lineární (liší se jen v matematických modelech v pozadí, jejichž znalost ale není pro praxi nutná). Dále se naučíme zahrnovat do naší analýzy i nejistoty měření našich hodnot (zejména nejistoty typu A způsobované použitým postupem měření). Nakonec stručně zmíníme, jak postupovat při testech hypotéz o hodnotách regresních koeficientů.

Nelineární regrese

Až do teď jsme se zabývali pouze případem, kdy prokládaná funkce byla lineární v neznámých regresních koeficientech, prokládaná funkce tedy měla vždy podobu

$$f(x) = \beta_0 + \beta_1 f_1(x) + \dots + \beta_k f_k(x).$$

Toto byla tzv. lineární regrese. Nyní se budeme zabývat případem, kdy potřebujeme naměřenými daty proložit funkci, která není lineární v neznámých regresních koeficientech. Budeme tedy prokládat funkci tvaru

$$f(x, \beta_0, \beta_1, \dots, \beta_k),$$

kde $\beta_0, \beta_1, \dots, \beta_k$ jsou neznámé regresní koeficienty, v nichž není tato funkce lineární. Takovémuto typu regrese se říká nelineární regrese. Naším cílem nadále bude stejně jako v případě regrese lineární odhadnout z naměřených dat hodnoty neznámých regresních koeficientů a vykreslit do grafu proloženou funkci.

Nelineární regrese se od lineární regrese liší jen v několika detailech. Na odhadování neznámých parametrů v nelineární regresí použijeme stejně jako v případě regrese lineární metodu maximální věrohodnosti, která za předpokladu normálního rozdělení chyb měření bude ekvivalentní metodě nejmenších čtverců. Výpočetní aspekt odhadu parametrů v nelineární regresí je značně komplikovanější a musí proto používat numerické algoritmy (na rozdíl od lineární regrese, kde existoval explicitní vzorec). Vlastnosti odhadů budou v případě nelineární regrese velice podobné regresí lineární. Všechny tyto odlišnosti si nyní popíšeme podrobněji.

Metoda maximální věrohodnosti

O principu fungování metody maximální věrohodnosti už jsme psali v minulém dílu seriálu, ale pro jistotu ji zde ještě stručně zopakujeme. Při metodě maximální věrohodnosti předpokládáme, že naměřená data (tedy dvojice (x_i, y_i)) byla vygenerována pomocí následujícího schématu. Pro každou hodnotu nezávisle proměnné x_i jsme naměřili určitou hodnotu y_i podle vztahu

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

kde ε_i představuje náhodnou nepřesnost měření, o které předpokládáme, že má rozdělení $N(0, \sigma^2)$, pro nějakou neznámou hodnotu σ^2 . Věrohodnost pro určitou volbu hodnot regresních koeficientů β_0, \dots, β_k je definována jako pravděpodobnost, že při takovéto volbě regresních koeficientů naměříme právě taková data, která jsme naměřili. Metoda maximální věrohodnosti říká, že za odhady regresních koeficientů vezmeme takové hodnoty $\hat{\beta}_0, \dots, \hat{\beta}_k$, které maximalizují věrohodnost pro naše naměřená data. Věrohodnost pro naše naměřená data je tvaru

$$L(\beta_0, \dots, \beta_k, x_1, \dots, x_n, y_1, \dots, y_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(y_i - f(x_i, \beta_0, \dots, \beta_k))^2}{\sigma^2}}.$$

Tento výraz se dá pomocí algebraických úprav přepsat na výraz

$$\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - f(x_i, \beta_0, \dots, \beta_k))^2}.$$

Podobně jako v minulém díle seriálu můžeme odvodit, že tento výraz je maximalizován, právě když je minimalizována hodnota výrazu

$$\sum_{i=1}^n (y_i - f(x_i, \beta_0, \dots, \beta_k))^2.$$

Chceme tedy za odhady regresních koeficientů zvolit takové hodnoty $\hat{\beta}_0, \dots, \hat{\beta}_k$, aby byl minimalizován součet čtverců vzdáleností naměřených hodnot od položené funkce. Používáme tedy metodu nejmenších čtverců.

Výpočetní aspekty odhadů parametrů

Jak už jsme naznačili, počítání odhadů regresních koeficientů v nelineární regresním modelu je značně komplikované. V praxi se pro tento výpočet musejí používat numerické metody. Nejčastěji používanou metodou je tzv. gradient descent metoda, kterou si zde stručně popíšeme.

Tato metoda se snaží najít extrém funkce tím způsobem, že začne v nějakém námi zadaném bodě (později bude upřesněno, jak zvolit počáteční bod), následně určí směr, ve kterém minimalizovaná funkce co nejvíce klesá³ a posune se v tomto směru o předem zvolený krok. Tento postup se pořád opakuje, čímž se postupně posouváme k hledanému minimu funkce. Pokud se dostaneme k hledanému minimu funkce, začne se tato metoda většinou cyklit na místě (už se nikam systematicky neposouvá). Když se toto stane, je to signál, že můžeme skončit a prohlásit poslední bod za aproximaci hledaného minima.

Tato metoda požaduje, aby uživatel na začátku zadal hodnotu počátečního bodu. Ve většině případů na volbě počátečního bodu nezáleží (tj. můžeme zadat libovolný počáteční bod), ale dobrá volba počátečního bodu dokáže zejména u složitějších úloh značně urychlit výpočet. Jako počáteční bod by se měl volit bod, o kterém si myslíme, že je blízko hledanému řešení (aby nebyla cesta k hledanému minimu příliš dlouhá). Jak už bylo řečeno dříve, špatná volba počátečního bodu obvykle nemá fatální důsledky (kromě delší doby výpočtu), ale může se stát, že kvůli špatné volbě počátečního bodu metoda nebude vracet požadované výsledky.

³Toto se provede pomocí gradientu funkce (odtud název gradient descent), což je něco jako zobecněná derivace pro funkce více proměnných.

Toto je nejčastěji způsobeno tím, že metoda nenajde globální minimum, ale pouze nějaké lokální minimum. Je proto dobré věnovat pozornost vhodné volbě startovacího bodu a potom také zkontrolovat, zda výstup programu odpovídá intuitivní představě o správném výsledku. V opačném případě bychom měli zkusit jinou volbu počátečního bodu. V příloženém vzorovém skriptu je opět uvedena ukázka toho, jak správně volit hodnoty počátečních bodů, abychom se vyhnuli nesprávným závěrům.

Pokud budeme chtít v programu R odhadovat regresní koeficienty v nelineární regresí, po zavolání příslušného příkazu na pozadí proběhne podobný algoritmus, jaký byl popsán výše. Z tohoto důvodu je tedy potřeba společně s ostatními vstupními parametry modelu zadat i počáteční bod pro metodu gradient descent. Nyní máme alespoň základní představu, proč se po nás něco takového chce.

Vlastnosti odhadů

V případě lineární regrese jsme odvodili poměrně dost vlastností, které mají odhady regresních koeficientů, a naučili jsme se konstruovat intervalové odhady pro hodnoty regresních koeficientů i prokládané funkce (viz minulý díl seriálu). Nyní bychom chtěli něco podobného odvodit i pro případ nelineární regrese.

Dobrá zpráva je, že v případě nelineární regrese mají naše odhady úplně analogické vlastnosti jako odhady v případě lineární regrese a dají se konstruovat také intervalové odhady pro regresní koeficienty i pro prokládanou funkci. Špatná zpráva je, že matematická teorie, která všechno toto dovoluje, je v případě nelineární regrese o dost složitější. Na tomto místě tedy tuto teorii nebudeme uvádět a jen napíšeme, že se dají zkonstruovat asymptotické intervalové odhady, které mají pro dostatečně velký počet měření (bude upřesněno dále) vlastnosti

$$P\left(\beta_i \in \left(\widehat{\beta}_i \pm u_{1-\frac{\alpha}{2}} s_n^{K_i}\right)\right) \doteq 1 - \alpha,$$

$$P\left(f(x, \beta_0, \dots, \beta_k) \in \left(f(x, \widehat{\beta}_0, \dots, \widehat{\beta}_k) \pm u_{1-\frac{\alpha}{2}} s_n^f(x)\right)\right) \doteq 1 - \alpha,$$

kde $s_n^{K_i}$ je nejistota měření regresního koeficientu a $s_n^f(x)$ je nejistota měření funkční hodnoty prokládané funkce v bodě x . Obě tyto hodnoty poskytuje matematický software jako standardní výstup. Jak už bylo zmíněno dříve, na výpočet těchto nejistot sice existují explicitní vzorce, my je zde ale nebudeme uvádět.

Regresní diagnostika

I v případě nelineární regrese je potřeba, aby byly splněny všechny předpoklady modelu, jinak obdržené výsledky nebudou správné. Předpoklady jsou v podstatě stejné jako v případě lineární regrese, pro jistotu je zde ale zopakujeme. Předpoklady nelineární regrese jsou

- Správná volba prokládané funkce.
- Stejný rozptyl chyb měření.
- Nezávislost jednotlivých měření.
- Normalita chyb měření.

Stejně jako v lineární regresi platí, že čtvrtý předpoklad není nijak zásadní, pokud pracujeme s dostatečným počtem měření. Splnění ostatních předpokladů je ale poměrně zásadní. Je proto nutné pokaždé provést alespoň základní regresní diagnostiku, abychom ověřili, že byly všechny předpoklady splněny (nebo že alespoň nebyly porušeny zásadním způsobem).

V zásadě platí, že na ověření platnosti předpokladů můžeme použít stejné metody, které používáme na ověření předpokladů v lineární regresi. Nemělo by smysl všechny tyto metody zde znovu opakovat, proto jen odkážeme na kapitulu Regresní diagnostika v minulém dílu seriálu.

Statistické testy o hodnotách regresních koeficientů

Metody popsané v tomto odstavci jsou aplikovatelné na lineární i nelineární regresi. Poslední věc, kterou bychom chtěli v souvislosti s regresní analýzou pokrýt, je testování hypotéz o hodnotách regresních parametrů. Je to tedy situace, kdy chceme pomocí naměřených dat otestovat hypotézu proti alternativě tvaru

$$\begin{aligned} H : \beta_j &= \vartheta, \\ A : \beta_j &\neq \vartheta, \end{aligned}$$

kde ϑ je nějaká předem zvolená konstanta.

Jak bylo popsáno ve čtvrtém dílu seriálu, k odvození testu potřebujeme odvodit podobu testové statistiky a kritického oboru. Jako testovou statistiku v tomto případě zvolíme

$$T = \frac{\hat{\beta}_j - \vartheta}{s_n^{K_j}},$$

kde $s_n^{K_j}$ je nejistota měření regresního koeficientu. Z předchozího dílu seriálu víme, že v případě lineární regrese je tato nejistota rovna prvku na pozici (j, j) v matici $\hat{\sigma}^2(\mathbb{X}^T\mathbb{X})^{-1}$, kde \mathbb{X} je matice modelu. V případě nelineární regrese jsme si explicitní vzorec neuvedli a spoléháme se v tomto na matematický software. Takto zvolená testová statistika bude za platnosti hypotézy konvergovat v distribuci k rozdělení $N(0, 1)$. Kritický obor tedy zvolíme jako

$$C = (-\infty, u_{\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}, \infty).$$

Zdůvodnění je stejné jako ve čtvrtém dílu seriálu u klasického t -testu.

Jednoduchou modifikací by se dal takovýto test upravit, aby testoval mírně pozměněnou hypotézu proti alternativě

$$\begin{aligned} H : \beta_j &< \vartheta, \\ A : \beta_j &\geq \vartheta. \end{aligned}$$

Testová statistika by se zvolila stejně, jen volba kritického oboru by se lišila. Kritický obor testu by v tomto případě byl

$$C = (u_{1-\alpha}, \infty).$$

Zdůvodnění je opět zcela analogické jako v případě t -testu. V případě opačných znamének nerovnosti v definici testované hypotézy a alternativy bychom dostali stejným způsobem kritický obor tvaru

$$C = (-\infty, u_{\alpha}).$$

Několik poznámek k nelineární regresi

- Stejně jako v případě lineární regrese platí, že v každém případě musíme mít alespoň takový počet měření, jako máme neznámých regresních koeficientů (jinak nemůžeme nelineární regresi ani použít). Dále platí, že aby byly intervalové odhady pro regresní koeficienty a hodnoty prokládané funkce, které jsme popsali v předchozích kapitolách, přesné, potřebujeme mít alespoň 4-5 krát více měření než regresních koeficientů.
- Stejně jako v případě lineární regrese platí, že čím více naměřených hodnot použijeme pro odhad regresních koeficientů, tím přesnější výsledky dostaneme (dostaneme menší nejistotu měření regresních parametrů).
- Stejně jako v případě lineární regrese platí, že bychom měli naměřenými daty prokládat jen takové funkce, které mají určité fyzikální opodstatnění. Z opačným případě se vystavujeme velkému riziku, že zvolíme prokládanou funkci špatně, což by mělo na závěry naší analýzy fatální důsledky.
- Další věc, na kterou si musíme dát v případě nelineární regrese pozor, je identifikovatelnost regresních parametrů v našem modelu (v lineární regresi jsme toto řešit nemuseli). Identifikovatelnost regresních parametrů znamená, že se nesmí stát, že bychom pro různé hodnoty regresních koeficientů dostali stejnou prokládanou funkci.

Pro lepší pochopení uvedeme příklad. Pokud bychom zvolili prokládanou funkci ve tvaru

$$f(x) = a + \frac{b}{cx + d},$$

stalo by se, že by regresní parametry nebyly identifikovatelné. Například pro volby parametrů

$$\begin{aligned}(a, b, c, d) &= (1, 1, 1, 1), \\ (a', b', c', d') &= (1, 2, 2, 2)\end{aligned}$$

bychom dostali naprosto stejnou prokládanou funkci (po krácení ve zlomku). Tento problém se dá vždy odstranit vhodnější parametrizací (s méně regresními koeficienty). V tomto případě bychom si museli prokládanou funkci přepsat do tvaru

$$f(x) = a + \frac{1}{\frac{c}{b}x + \frac{d}{b}} = \beta_0 + \frac{1}{\beta_1 x + \beta_2},$$

kde máme jen 3 regresní koeficienty $\beta_0, \beta_1, \beta_2$.

- Pokud používáme pro zpracování měření určitého experimentu nelineární regresi, měli bychom do závěrečného protokolu uvést dostatečně mnoho informací, aby čtenář byl schopen přesně zrekonstruovat náš postup, poskytnout i grafické výstupy a dbát přitom na přehlednost. Stejně jako v případě aplikace lineární regrese platí, že bychom jako výstup měli uvést minimálně
 - Tvar prokládané funkce (tedy vzorec $f(x) = \dots$)
 - Bodový odhad a nejistota měření všech regresních koeficientů.

- Alespoň stručný komentář, zda jsou splněny všechny předpoklady použití regresního modelu (případně upozornění na možné nepřesnosti způsobené nesplněním předpokladů). Není nutné přikládat všechny popsané grafy.
- Pokud se rozhodnete vykreslit graf s naměřenými daty a proloženou funkcí, měl by být v legendě uveden tvar prokládané funkce. Je také dobré (i když ne úplně nutné) do grafu vykreslit interval spolehlivosti pro prokládanou funkci. Ve vzorovém skriptu najdete podrobný návod a ukázkou, jak by toto mělo správně vypadat.
- Pozorného čtenáře jistě napadne, že nelineární regrese je univerzálnější než lineární regrese a že bychom mohli používat pouze nelineární regresi (tedy aplikovat ji i na problémy řešitelné lineární regresí). Toto je do určité míry pravda, ale stále existuje mnoho důvodů, proč je dobré používat lineární regresi v případech, kdy to jde. Mezi hlavní patří, že lineární bude poskytovat přesnější výsledky (díky jednodušší formulovaným modelům a jednodušší matematické teorii). Jako druhý hlavní důvod můžeme uvést snadnější výpočet koeficientů. V problémech, které v tomto seriálu řešíme, a při uvážení výpočetní síly dnešních počítačů se toto nezdá být jako nevýhoda nelineární regrese, ale ocenili bychom to v případě zpracovávání opravdu velkých objemů měřených dat.

Pokročilé partie regresní analýzy

V této poslední kapitole se budeme věnovat několika problémům pokročilé regresní analýzy. Uvádíme je zde spíše pro zajímavost, neboť na ně v praxi moc často nenarazíme a dá se říci, že je v určitém přiblížení můžeme řešit už známými metodami nebo tyto problémy ignorovat a nedopustili bychom se velkých nepřesností (alespoň tedy ve velké většině případů).

Omezení na regresní koeficienty

Někdy v praxi potkáme případ, kdy předem víme, že regresní koeficienty musí nabývat jen určitých hodnot. Například mohou nabývat jen kladných hodnot nebo jen "malých hodnot" (řekněme menší než 100). V takovémto případě je možno numerickému algoritmu tuto informaci poskytnout a ten bude hledat regresní koeficienty tak, aby tyto naše podmínky splňovaly. V příloženém vzorovém skriptu najdete příklad, kdy je toto potřeba udělat, a také, jak to správně udělat.

Toto se nejčastěji stane, když prokládáme naměřenými daty periodické funkce (jako např. funkce $\sin(x)$, $\cos(x)$ atd.). Představme si, že chceme naměřenými daty proložit funkci

$$f(x) = a + b \sin(cx + d).$$

Pokud bychom si nestanovili žádnou dodatečnou podmínkou na hodnoty regresních koeficientů, velmi pravděpodobně by nám software na výstupu poskytl velice vysokou hodnotu odhadu regresního koeficientu c . Pokud bychom si následně nakreslili graf proložené funkce, dostali bychom sinusoidu s velkou frekvencí, která by sice skoro dokonale prošla naměřenými daty (tedy součet čtverců residuí by byl velice malý), ale velice špatně by aproximovala pozorovanou závislost. Řešením by v tomto případě bylo omezit možné hodnoty regresního koeficientu c . Příklad správného použití omezujících podmínek najdete ve vzorovém skriptu.

Problematická volba prokládané funkce - spliny

V praxi se může někdy stát, že závislost, kterou chceme naměřenými daty proložit, je příliš komplikovaná na to, abychom ji mohli vyjádřit pomocí jedné prokládané funkce s regresními koeficienty. Pokud bychom chtěli i v této situaci vykreslit graf, ve kterém bude kromě naměřených dat i proložená teoretická křivka, můžeme k tomu použít lokální aproximaci polynomy.

Lokální aproximace funkce polynomy zjednodušeně znamená, že definiční obor funkce rozdělíme na menší intervaly a na každém takovémto intervalu budeme chtít najít polynom, který bude co nejvíce podobný aproximované funkci (podobný ve smyslu, že vzdálenost mezi aproximovanou funkcí a polynomem bude co nejmenší). Jistě někteří z vás už slyšeli o Taylorových polynomech. Taylorovy polynomy jsou jedním ze způsobů, jak lokálně aproximovat funkci pomocí polynomů⁴.

Idea, jak budeme chtít postupovat, je taková, že si naměřená data rozdělíme do několika intervalů podle hodnot nezávisle proměnné (dělicí body budeme v tomto kontextu nazývat uzly) a na každém z těchto intervalů budeme chtít naměřenými daty prokládat polynomiální funkci určitého stupně p . Jednotlivým polynomům se v tomto kontextu říká regresní spliny⁵. V praxi stačí zvolit p rovno dvěma nebo třem, větší stupně prokládaných polynomů už obvykle nevedou k výrazně lepším výsledkům.

K proložení polynomů na jednotlivých intervalech můžeme použít lineární regresi, která byla popsána v minulém dílu seriálu. Regresní koeficienty budeme tedy odhadovat pomocí metody nejmenších čtverců. Pokud bychom vše provedli jen tak, jak jsme dosud popsali, velice pravděpodobně bychom dostali nespojitou proloženou funkci, což není úplně to, co by odpovídalo intuitivním požadavkům na prokládanou funkci. Proto si navíc přidáme podmínku na to, aby bylo napojení proložených polynomů na krajích našich intervalů spojitě. Takovouto podmínku přidáme jednoduše tím, že při počítání odhadů regresních koeficientů budeme uvažovat jen taková řešení, která zároveň splňují podmínku na spojitost na krajích intervalů. Pokud si nejste jistí, jak by se takováto soustava řešila, nevádí, v praxi toto vždy bude obstarávat matematický software. Tento odstavec slouží jen na získání základní představy, co se v počítači po spuštění příkazu děje.

Poslední problém, který musíme vyřešit, je volba dělení na intervaly, tedy jaké množství a jaké rozmístění uzlů zvolit. Obecně se dá říci, že je dobré zvolit uzly tak, aby na každém vzniklém intervalu stále byl dostatečný počet měření (ideálně alespoň 4 až 5 krát více, než je stupeň prokládaného polynomu). Uzly bychom měli potom volit tak, aby hranice intervalů odpovídaly bodům, ve kterých se mění chování měřených dat. Obvykle se v praxi postupuje metodou pokus-omyl, tedy zkusíme podle výše popsaných pravidel uzly nějak zvolit, vykreslíme si proloženou funkci a pokud nejsme spokojeni, pokoušíme se jinou volbou uzlů výsledek zlepšit. Zejména bychom volbou uzlů měli zajistit, aby proložená funkce nevykazovala příliš rozkolísané a neočekávané chování. Tato metoda vyžaduje cvik a zkušenosti, proto ve vzorovém skriptu najdete příklady toho, jak by správná volba uzlů měla a neměla vypadat.

Pomocí tohoto postupu dostaneme pěkně vypadající graf, ale je nutné si uvědomit několik omezení, která tento postup má:

- Je potřeba poměrně hodně měření k tomu, aby byla tato metoda spolehlivá (což je logické vzhledem k tomu, že naměřená data rozdělíme na několik skupin a na nich zvlášť prokládáme jednotlivé funkce).

⁴Pokud jste o Taylorových polynomech ještě neslyšeli, nevádí, uvádíme to zde jen pro zajímavost.

⁵Čti [splajny].

- Obdržené odhady regresních koeficientů na rozdíl od klasické lineární regrese nemají na-prosto žádnou fyzikální interpretaci. Jediné, čeho tímto postupem docílíme, je názorně proložená křivka v grafu naměřených hodnot.
- Vystává problém, jaké dělení na jednotlivé intervaly bychom měli správně zvolit, neboť obdržený výsledek závisí na zvoleném dělení.

Aplikace regrese v případě heteroskedasticity

V tomto dílu a v minulém dílu seriálu jsme opakovaně uváděli, že jedním z předpokladů lineárních i nelineárních regresních modelů je shodnost rozptylu měřených dat (homoskedasticita). Uvedli jsme, že pokud není tento předpoklad porušen výrazně, můžeme toto porušení ignorovat (případně slovně upozornit na možné nepřesnosti v obdržených výsledcích). Nyní představíme metodu, jak můžeme heteroskedasticitu vzít v potaz při odhadování modelu tak, abychom obdrželi správné výsledky.

Pro začátek je nutné si uvědomit, že odhady regresních koeficientů budou správné i při porušení předpokladu homoskedasticity, jediné, co bude tímto předpokladem pokazeno, jsou nejistoty odhadů regresních koeficientů. Tyto nejistoty odhadů budou ve většině případů podhodnoceny, budeme tedy mít falešný pocit, že jsme koeficienty změřili přesně, i když opak je ve skutečnosti pravdou. Naštěstí existuje metoda, jak upravit nejistoty měření regresních koeficientů tak, aby reflektovaly nesplněný předpoklad homoskedasticity. Této metodě se říká Whiteův odhad kovarianční matice (někdy také sendvičový odhad). Je nad možností tohoto seriálu podrobně popsat tento odhad, proto to na tomto místě ani nebudeme dělat. Jediné, co je potřeba si pamatovat, je, že pokud pomocí regresní diagnostiky odhalíme nesplnění předpokladu homoskedasticity, je potřeba použít jiný výpočet nejistot měření, než je ten základní. Ve vzorovém skriptu můžete nalézt příklad, jak takovýto odhad v praxi konstruovat.

Typickým důsledkem použití Whiteova odhadu kovarianční matice namísto základního odhadu kovarianční matice je zvětšení nejistot odhadů regresních parametrů. Toto je nepříjemné z hlediska snížení přesnosti výsledků, ale je nutné si uvědomit, že je to jediná možnost, neboť předchozí nejistoty měření regresních koeficientů byly sice menší, ale nebyly správné. Přítomnost heteroskedasticity v našich datech snižuje přesnost, se kterou jsme schopni určit regresní koeficienty.

Pozorného čtenáře určitě napadne, proč bychom nemohli Whiteův odhad kovarianční matice používat vždy (je to přece obecnější odhad). Toto by bylo možné a obdržené výsledky by byly správné, ale všechny naše výpočty by se výrazně zkomplikovaly. Navíc je nutné zmínit, že mírné porušení předpokladu homoskedasticity příliš nevadí (nijak výrazně to neovlivní správnost získaných výsledků) a můžeme ho ignorovat (jak bylo zmiňováno dříve). Whiteův odhad je nutné používat až v případech opravdu vážného porušení předpokladu homoskedasticity. Ve vzorovém skriptu opět můžete najít ukázky použití Whiteova odhadu včetně ukázky toho, kdy už je nutné tento odhad používat a kdy to ještě není nutné.

Regrese na základě průměrů z měření

Pokud se provádí nějaké rozsáhlejší experimentální měření, většinou se postupuje tak, že se pro každou hodnotu nezávisle proměnné změní více hodnot závisle proměnné⁶. Matematická teorie

⁶Potom se (mimo jiné výhody) dá provádět test vhodnosti prokládané funkce popsany v minulém díle seriálu.

regresní analýzy potom umožňuje dělat regresi jen na základě průměru naměřených hodnot pro jednotlivé volby nezávisle proměnné.

Pro lepší představu si uvedeme příklad. Představme si, že naměříme následující data

$$\begin{aligned} & (x_1, y_{1,1}), \dots, (x_1, y_{1,n_1}), \\ & \vdots \\ & (x_m, y_{m,1}), \dots, (x_m, y_{m,n_m}). \end{aligned}$$

Takováto situace není nic neobvyklého, na základě dosud vyložené teorie si s ní umíme poradit (tj. umíme těmito daty proložit libovolnou funkční závislost). Co chce tento odstavec říci, je, že pokud by nám někdo dal jen částečnou informaci o těchto měřeních, a to sice průměry hodnot závisle proměnné v jednotlivých skupinách odpovídajících hodnotám nezávisle proměnné, příslušnou výběrovou směrodatnou odchylku průměru a počty měření v jednotlivých skupinách, tedy

$$(x_1, \overline{y_{1,\bullet}}, s_{n_1}, n_1), \dots, (x_m, \overline{y_{m,\bullet}}, s_{n_m}, n_m),$$

stále bychom byli schopni takovými daty proložit libovolnou funkční závislost. Někdy se místo surových dat poskytují právě jen průměry hodnot nezávisle proměnné, příslušné výběrové směrodatné odchylky průměrů a počty měření v jednotlivých skupinách z důvodu úspory času a paměťové náročnosti.

Toto tvrzení si nyní také odvodíme. Budeme postupovat tak, že si napíšeme součet čtverců, který chceme při odhadování neznámých regresních koeficientů minimalizovat, na základě všech měření (tedy zatím nebudeme zavádět zjednodušení měřených dat). Následně pomocí algebraických úprav ukážeme, že jsme schopni tento součet čtverců přepsat jen pomocí symbolů $(x_j, \overline{y_{j,\bullet}}, s_{n_j}, n_j)$, čímž bude odvození dokončeno, neboť při praktickém výpočtu můžeme používat jen tento alternativní zápis. Nyní už k samotnému odvození, součet čtverců, který chceme minimalizovat, má tvar

$$S(x_1, \dots, x_j, y_{1,1}, \dots, y_{j,n_j}, \beta_0, \dots, \beta_k) = \sum_{i=1}^N (y_i - f(x_j, \beta_0, \dots, \beta_k))^2,$$

kde N představuje celkový počet měření. Tento výraz budeme nyní algebraicky upravovat, nejprve rozdělíme sumu přes všechna měření na dvě sumy, nejprve podle hodnot nezávisle proměnné x a poté podle hodnot měření příslušných měření závisle proměnné, čímž dostaneme

$$\sum_{j=1}^m \sum_{i=1}^{n_j} (y_i - f(x_j, \beta_0, \dots, \beta_k))^2.$$

Nyní provedeme důležitý trik, a sice ten, že k vnitřku závorky přičteme a odečteme výběrový průměr příslušné skupiny měření, tedy člen

$$\overline{y_{j,\bullet}} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{j,i}.$$

Tímto nezměníme hodnotu našeho výrazu, ale následně si tím výrazně pomůžeme, dostáváme

$$\sum_{j=1}^m \sum_{i=1}^{n_j} (y_{j,i} - \overline{y_{j,\bullet}} + \overline{y_{j,\bullet}} - f(x_j, \beta_0, \dots, \beta_k))^2.$$

Nyní roznásobíme závorku, čímž dostaneme

$$\sum_{j=1}^m \sum_{i=1}^{n_j} \left[(y_{j,i} - \overline{y_{j,\bullet}})^2 + 2(y_{j,i} - \overline{y_{j,\bullet}})(\overline{y_{j,\bullet}} - f(x_j, \beta_0, \dots, \beta_k)) + (\overline{y_{j,\bullet}} - f(x_j, \beta_0, \dots, \beta_k))^2 \right].$$

Nyní tento výraz rozdělíme na jednotlivé sumy a uvědomíme si, že některé vzniklé výrazy můžeme z definice přepsat pomocí výrazů $s_{n_i}^2$ a n_i , dostáváme

$$\begin{aligned} & \sum_{j=1}^m \sum_{i=1}^{n_j} (y_{j,i} - \overline{y_{j,\bullet}})^2 + \sum_{j=1}^m \sum_{i=1}^{n_j} (\overline{y_{j,\bullet}} - f(x_j, \beta_0, \dots, \beta_k))^2 + \\ & + \sum_{j=1}^m \sum_{i=1}^{n_j} 2(y_{j,i} - \overline{y_{j,\bullet}})(\overline{y_{j,\bullet}} - f(x_j, \beta_0, \dots, \beta_k)) = \\ & = \sum_{j=1}^m (n_j - 1)s_{n_j}^2 + \sum_{j=1}^m n_j (\overline{y_{j,\bullet}} - f(x_j, \beta_0, \dots, \beta_k))^2 + \\ & + \sum_{j=1}^m \sum_{i=1}^{n_j} 2(y_{j,i} - \overline{y_{j,\bullet}})(\overline{y_{j,\bullet}} - f(x_j, \beta_0, \dots, \beta_k)). \end{aligned}$$

První dva členy našeho součtu už jsou vyjádřeny jen pomocí symbolů $(x_i, \overline{y_{j,\bullet}}, s_{n_i}, n_i)$, zbývá nějak se vypořádat se třetím členem. Na tomto místě si musíme uvědomit, že třetí člen je roven nule, což plyne z jednoduchých algebraických úprav, které zde nyní podrobně provedeme.

$$\begin{aligned} & \sum_{j=1}^m \sum_{i=1}^{n_j} 2(y_{j,i} - \overline{y_{j,\bullet}})(\overline{y_{j,\bullet}} - f(x_j, \beta_0, \dots, \beta_k)) = \\ & = 2 \sum_{j=1}^m \sum_{i=1}^{n_j} [y_i \overline{y_{j,\bullet}} - y_{j,i} f(x_j, \beta_0, \dots, \beta_k) - (\overline{y_{j,\bullet}})^2 + \overline{y_{j,\bullet}} f(x_j, \beta_0, \dots, \beta_k)] = \\ & = 2 \sum_{j=1}^m [n_j (\overline{y_{j,\bullet}})^2 - n_j \overline{y_{j,\bullet}} f(x_j, \beta_0, \dots, \beta_k) - n_j (\overline{y_{j,\bullet}})^2 + n_j \overline{y_{j,\bullet}} f(x_j, \beta_0, \dots, \beta_k)] = \\ & = 0. \end{aligned}$$

Shrneme si, co jsme celkově dokázali. Zapsali jsme součet čtverců, který chceme při odhadování neznámých regresních koeficientů minimalizovat, pouze pomocí symbolů $(x_i, \overline{y_{i,\bullet}}, s_{n_i}, n_i)$ jako

$$S(x_1, \dots, x_j, y_{1,1}, \dots, y_{j,n_j}, \beta_0, \dots, \beta_k) = \sum_{j=1}^m (n_j - 1)s_{n_j}^2 + \sum_{j=1}^m n_j (\overline{y_{j,\bullet}} - f(x_j, \beta_0, \dots, \beta_k))^2.$$

Je tedy vidět, že k určení odhadů regresních koeficientů nepotřebujeme znát hodnoty všech měření, ale stačí nám znát hodnoty $(x_j, \overline{y_{j,\bullet}}, s_{n_j}, n_j)$. V praxi se potom na hledání odhadů regresních koeficientů použijí stejné metody, které byly popsány dříve, a bude to za nás vždy provádět počítač.

Když se do takovéto situace dostaneme a chceme vykreslit graf naměřených hodnot a proložené funkce, většinou do grafu nekreslíme samotné naměřené hodnoty, ale pouze průměry měření odpovídajících hodnotám nezávisle proměnné $\overline{y_{j,\bullet}}$ a nejistoty s_{n_j} , které vyjádříme pomocí tzv. error barů (česky chybová úsečka). Chybová úsečka je vertikální úsečka o velikosti $2s_{n_j}$ se středem v bodě $y_{j,\bullet}$, pomocí které v některých případech lépe vyjádříme rozptýlenost měřených hodnot (zejména v případech, kdy máme hodně naměřených hodnot a klasický graf by byl kvůli tomu nepřehledný). Delší chybová úsečka znamená, že měřené hodnoty byly více rozptýleny. Chybová úsečka se dá také interpretovat tak, že je pravděpodobnost přibližně 68 %, že skutečná hodnota prokládané funkce leží uvnitř chybové úsečky (pokud máme dostatek měření, abychom mohli použít centrální limitní větu). V příloženém vzorovém skriptu také naleznete příklad na použití této metody a kreslení chybových úseček.

Regrese při uvažování nepřesností měření nezávisle proměnné

Až dosud jsme uvažovali pouze nepřesnosti měřené závisle proměnné y , hodnoty nezávisle proměnné x byly považovány za přesně známé. Nyní si rozebereme obecnější případ, kdy uvažujeme i hodnoty nezávisle proměnné x za zatížené nepřesnostmi měření. Naším cílem bude opět prokládat naměřenými daty teoretickou závislost, k tomu budeme chtít odhadovat neznámé regresní koeficienty z naměřených dat.

Stejně jako v předchozích případech se na odhad hodnot regresních koeficientů použije metoda maximální věrohodnosti. Nyní ale bude analytické vyjádření řešeného problému o něco složitější. Obecně budeme předpokládat, že pro měřená data platí teoretická závislost

$$y = f(x, \beta_0, \dots, \beta_k),$$

ale my nemůžeme přímo měřit hodnoty x a y . Tyto veličiny budeme vždy měřit nepřesně, výsledné hodnoty měření \tilde{x}_i, \tilde{y}_i tedy budou rovny

$$\begin{aligned}\tilde{y}_i &= f(x_i, \beta_0, \dots, \beta_k) + \varepsilon_i, \\ \tilde{x}_i &= x_i + \delta_i,\end{aligned}$$

kde ε_i je náhodná veličina s rozdělením $N(0, \sigma_y^2)$ a δ_i je náhodná veličina s rozdělením $N(0, \sigma_x^2)$.

Nebudeme na tomto místě podrobně odvozovat, jak budou odhady v takovémto případě vypadat, neboť je to příliš komplikované a v praxi je za nás bude vždy dělat počítač. Jediné, co zde uvedeme, je, jak si můžeme takovouto metodu představit ve srovnání s klasickou regresí při uvažování měření hodnot nezávisle proměnné bez nepřesností. V případě klasické regrese jsme chtěli minimalizovat součet druhých mocnin vertikálních vzdáleností naměřených hodnot od prokládané funkce. Vertikální vzdálenost naměřené hodnoty od prokládané funkce byla nazývána residuum, matematicky zapsáno

$$U_i = y_i - f(x_i, \beta_0, \dots, \beta_k).$$

V případě, kdy uvažujeme i nepřesnosti měření nezávisle proměnné, ale nebudeme uvažovat vertikální vzdálenost, ale přímo vzdálenost naměřené hodnoty od prokládané funkce. Abychom mohli tuto vzdálenost správně upravit, potřebujeme znát poměr mezi rozptyly náhodných nepřesností měření příslušných nezávisle a závisle proměnné, tedy poměr

$$\frac{\sigma_x^2}{\sigma_y^2}.$$

Na závěr poznamenejme, že v praxi nastane téměř vždy případ, kdy měříme i hodnoty nezávisle proměnné s nepřesnostmi (minimálně nejistota typu B způsobená použitím nedokonalého měřidla). Téměř nikdy se nestane, že bychom hodnoty nezávisle proměnné znali přesně. Pokud ale je rozptyl hodnot měření nezávisle proměnné σ_x^2 výrazně menší, než rozptyl hodnot měření závisle proměnné, tato metoda dá téměř shodné výsledky jako klasická regrese (bez uvažování nepřesností nezávisle proměnné). Je to způsobeno tím, že úprava v chápání vzdálenosti nebude příliš odlišná od klasické vertikální vzdálenosti. V takovýchto případech je potom naprosto postačující použít metody klasické regrese, není nutné věci příliš komplikovat (výsledky budou jen zanedbatelně odlišné). Tato popsaná metoda by se měla aplikovat jen v případech, kdy jsou hodnoty σ_x^2 a σ_y^2 srovnatelné, což se v praxi často nestává.

Toto je způsobeno tím, že hodnoty nezávisle proměnné jsou typicky zatíženy jen nejistotou typu B (nepřesností měřidla), zatímco u hodnot závisle proměnné se musí počítat s nejistotami typu A (náhodnost měřených dat) i B (nepřesností měřidla). Toto způsobí, že hodnoty závisle proměnné mají řádově větší rozptyl než hodnoty nezávisle proměnné. Je ale nutné mít na paměti, že v případech, kdy jsou nepřesnosti měření závisle a nezávisle proměnné srovnatelné (což se také občas stává), metody klasické regrese jsou nedostačující a musíme proto použít výše popsaný zobecněný postup.

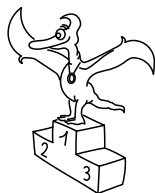
Jen úplně na závěr poznamenejme, že i v takovémto obecnějším případě lze konstruovat intervalové odhady pro hodnoty regresních koeficientů i pro hodnotu prokládané funkce. Teoretické odvození je však příliš složité na to, abychom ho zde uváděli. Ve vzorovém skriptu můžete najít příklady použití této metody na konkrétních datech.

Závěr

Toto je ze seriálu 30. ročníku FYKOSu všechno. Pokud jste dočetli až sem, musím Vám pogratulovat a zároveň poděkovat za trpělivost. Doufám, že znalosti, které jste v tomto seriálu načerpali, v budoucnosti budete moci využít, ať už v rámci různých fyzikálních soutěží nebo během studia na střední (nebo i vysoké) škole.

Na závěr bych chtěl ještě kromě čtenářů seriálu a řešitelů seriálových úloh poděkovat i korektorům, kteří se výrazným způsobem podíleli na finální podobě jednotlivých dílů tohoto seriálu.

Michal Nožička



Pořadí řešitelů po V. sérii



Kompletní výsledky najdete na <http://fykos.cz>.

Kategorie prvních ročníků

jméno	škola	1	2	3	4	5	P	E	S	V	%	Σ
<i>Student Pilný</i>	MFF UK	6	6	6	8	8	8	12	10	64	100	316
1. <i>Martin Vavřík</i>	G, Šumperk	6	6	6	2	6	6	12	12	56	84	264
2. <i>Václav Zvoníček</i>	G Brno, tř. Kpt. Jaroše	8	8	5	3	6	–	10	10	50	85	263
3. <i>Jakub Jobus</i>	G PdC, Piešťany	8	6	7	2	5	4	10	12	54	78	240
4. <i>Martin Schmied</i>	G Jihlava	8	6	7	4	4	4	10	12	55	75	232
5. <i>Radka Křížová</i>	G J. Heyrovského, Praha	6	6	2	–	5	8	–	3	30	72	200
6. <i>Viktor Materna</i>	G Brno, tř. Kpt. Jaroše	8	6	6	2	5	5	12	–	44	73	184
7. <i>Matěj Prokop</i>	G Dašická, Pardubice	4	4	2	–	5	–	–	10	25	79	174
8. <i>Jiří Zelenka</i>	G Z. Wintra, Rakovník	6	8	3	2	5	3	10	–	37	58	121
9. <i>Miroslav Macko</i>	ŠpMNDaG, Bratislava	8	–	6	–	–	3	–	–	17	72	113
10. <i>Jakub Pravda</i>	ŠpMNDaG, Bratislava	2	–	1	–	–	4	–	–	7	54	107
11. <i>Eva Vochozková</i>	Biskupské G, Brno	2	6	6	–	–	–	–	–	14	66	102
12. <i>Matěj Krátký</i>	PORG, Praha	8	6	–	–	6	–	–	3	23	80	90
13. <i>Šimon Brázda</i>	G, Hlinsko	6	–	–	–	–	–	–	–	6	56	82
14.–15. <i>Kateřina Barotová</i>	G, Olomouc-Hejčín	8	–	6	–	–	–	–	–	14	76	66
14.–15. <i>Pavla Rudolfová</i>	G, Vídeňská, Brno	2	6	6	–	–	7	–	–	21	72	66
16. <i>Richard Hamerlik</i>	ŠpMNDaG, Bratislava	–	–	–	–	–	–	–	–	–	45	65
17. <i>Petr Doubravský</i>	Akademické G, Praha	–	–	–	–	–	–	–	–	–	67	58
18. <i>Filip Novotný</i>	G Jihlava	–	–	–	–	–	–	–	–	–	63	48
19. <i>Adam Vavrečka</i>	G P. Bezruče, Frýdek-Místek	–	–	–	–	–	–	–	–	–	71	37
20. <i>Filip Wagner</i>	G Tišnov	–	–	–	–	–	–	–	–	–	61	34
21.–22. <i>Timea Szöllősová</i>	G Grösslingova, Bratislava	–	–	–	–	–	2	–	–	2	40	32
21.–22. <i>Jan Vondra</i>	G Týn nad Vltavou	4	–	0	–	–	–	–	–	4	39	32
23. <i>Lucia Gintnerová</i>	G Sv. Františka, Žilina	–	–	–	–	–	7	–	–	7	85	28
24. <i>Lukáš Hronek</i>	G, Písek	6	–	–	–	–	–	–	–	6	73	27
25. <i>Viktor Vařeka</i>	G P. Bezruče, Frýdek-Místek	–	–	–	–	–	–	–	–	–	76	25
26. <i>Marek Čermoch</i>	G F. Palackého, Val. Mez.	–	–	–	–	–	–	–	–	–	84	21
27. <i>Lucia Krajčoviechová</i>	G Jura Hronca, Bratislava	–	–	–	–	–	–	–	–	–	77	17
28. <i>Ján Šrejbr</i>	G J. Jungmanna, Litoměřice	–	–	–	–	–	–	–	–	–	42	16
29. <i>Marco Souza de Joode</i>	G Nad Štolou, Praha	2	6	6	1	–	–	–	–	15	58	15
30. <i>Vojtěch Ježek</i>	G Legionářů, Příbram	–	–	–	–	–	–	–	–	–	117	14
31. <i>Viktor Fukala</i>	G Jana Keplera, Praha	8	–	–	–	–	–	5	–	13	81	13
32.–33. <i>Soňa Curylová</i>	G F. Palackého, Val. Mez.	–	–	–	–	–	–	–	–	–	50	12
32.–33. <i>Milan Tichavský</i>	Slezské G, Opava	–	–	–	–	–	–	–	–	–	100	12
34.–35. <i>Anna Hollmannová</i>	SG Dr. Randy, Jablonec n. N.	–	–	–	–	–	–	–	–	–	50	10
34.–35. <i>Bohumír Žůrek</i>	G Volgogradská 6a, Ostrava	–	–	–	–	–	–	–	–	–	17	10
36. <i>Tomáš Salavec</i>	BG B. Balbína, Hradec Králové	–	–	–	–	–	–	–	–	–	73	8

Kategorie druhých ročníků

jméno	škola	1	2	3	4	5	P	E	S	V	%	Σ
<i>Student</i>	<i>Pilný</i>	MFF	UK									
		6	6	6	8	8	8	12	10	64	100	316
1. <i>Kateřina Rosická</i>	G J. Ortena, Kutná Hora	8	6	6	2	7	7	12	12	60	94	297
2. <i>Ladislav Trnka</i>	G, Havlíčkův Brod	8	6	7	3	6	7	10	12	59	89	275
3. <i>Jindřich Dušek</i>	G Jana Keplera, Praha	6	8	6	2	6	4	-	9	41	82	233
4. <i>Lucie Kundratová</i>	G, nám. TGM, Zlín	8	6	2	3	5	7	-	3	34	75	215
5. <i>Ivan Hudák</i>	ESS, Lip. Mikuláš	8	6	6	-	5	3	9	9	46	83	208
6. <i>Josef Minařík</i>	G Brno, tř. Kpt. Jaroše	6	6	-	-	-	-	-	-	12	81	206
7. <i>Jakub Růžička</i>	G, Nymburk	8	6	6	-	4	-	10	-	34	81	177
8. <i>Jiří Blaha</i>	G, Uherské Hradiště	6	6	6	-	-	-	-	-	18	86	142
9. <i>Křtstian Matušík</i>	G, Benešov	8	-	0	1	4	2	7	-	22	54	129
10. <i>Jakub Smolka</i>	Slezské G, Opava	6	6	6	-	-	-	-	-	18	61	93
11. <i>Marie Grunová</i>	G Moravský Krumlov	6	6	-	5	-	10	-	-	27	61	81
12. <i>Dominik Beňo</i>	G L. Svobodu, Humenné	8	8	-	-	-	10	-	-	26	100	78
13. <i>Marek Jankola</i>	G M. Hattalu, Trstená	6	6	5	-	4	-	6	3	30	73	71
14.-15. <i>Karel Balej</i>	G a SOŠ, Rokycany	-	-	-	-	-	-	-	-	-	95	59
14.-15. <i>Tereza Pavlišová</i>	G, Olomouc-Hejčín	-	-	-	-	-	-	-	-	-	63	59
16. <i>Kateřina Charvátová</i>	G B. Němcové, HK	6	4	-	-	-	-	-	-	10	65	53
17. <i>Veronika Vohníková</i>	Nový PORG, Praha	-	-	-	-	-	-	-	-	-	73	35
18. <i>Jakub Zemek</i>	G, Uherské Hradiště	-	-	-	-	-	-	-	-	-	61	28
19. <i>Jakub Rajnštajn</i>	G F. M. Pelcla, Rychnov n. Kn.	-	-	-	-	-	-	-	-	-	69	25
20.-22. <i>Martin Dinh</i>	G, Třinec	-	-	-	-	-	-	-	-	-	100	17
20.-22. <i>Daniel Martyněk</i>	G, Třinec	-	-	-	-	-	-	-	-	-	100	17
20.-22. <i>Štěpán Stryja</i>	G, Třinec	-	-	-	-	-	-	-	-	-	68	17
23. <i>Šimon Kondrk</i>	SPŠ Dubnica nad Váhom	-	-	-	-	-	-	-	-	-	67	16
24. <i>Andrea Bínová</i>	G, Česká Lípa	-	-	-	-	-	-	-	-	-	75	15
25.-26. <i>Josef Sabol</i>	G Chotěboř	-	-	-	-	-	-	-	-	-	100	12
25.-26. <i>Marie Vaňáková</i>	G Botičská, Praha	-	-	-	-	-	-	-	-	-	39	12
27.-28. <i>Máté Eke</i>	Gymnázium, sSOŠ a jazyková škola	-	-	-	-	-	-	-	-	-	58	11
27.-28. <i>Michal Jireš</i>	G F. M. Pelcla, Rychnov n. Kn.	-	-	-	-	-	-	-	-	-	85	11
29. <i>Klára Nechanická</i>	G Neumannova, Žďár n. S.	-	-	-	-	-	-	-	-	-	83	10
30. <i>Václav Bulín</i>	G, Plasy	-	-	-	-	-	-	-	-	-	82	9

Kategorie třetích ročníků

jméno	škola	1	2	3	4	5	P	E	S	V	%	Σ
<i>Student Pilný</i>	MFF UK	3	3	6	8	8	8	12	10	58	100	286
1. <i>Viktor Rosman</i>	G, Pelhřimov	4	4	6	4	7	7	12	–	44	90	239
2. <i>Tomáš Dulava</i>	Matičín G, Ostrava	3	3	6	2	5	5	4	9	37	72	205
3. <i>Vít Beran</i>	Masarykovo G, Plzeň	–	4	–	5	–	7	–	8	24	91	171
4. <i>Zuzana Richterová</i>	G, Pelhřimov	2	3	2	1	2	4	10	7	31	49	140
5.–6. <i>Martin Okáník</i>	G Tajovského, B. Bystrica	–	–	–	–	–	–	–	–	–	86	136
5.–6. <i>Dominik Starý</i>	G, Benešov	4	3	–	6	–	–	–	–	13	80	136
7. <i>David Košťák</i>	G, Josefská, Praha	–	–	–	–	–	–	–	–	–	71	114
8. <i>Katarína Častulíková</i>	1. súkromné G v Bratislave	2	3	0	–	5	2	8	–	20	58	113
9. <i>Ondřej Knopp</i>	G Christiana Dopplera, Praha	–	–	–	–	–	–	–	–	–	96	91
10. <i>David Dvořák</i>	ZŠ a G, Konice	–	–	–	–	–	–	–	–	–	50	54
11. <i>Daniela Hrbáčová</i>	Wichterlovo G, Ostrava	4	–	–	–	–	–	–	–	4	48	47
12.–13. <i>Ondřej Buček</i>	G Brno, tř. Kpt. Jaroše	4	3	5	–	–	–	–	–	12	65	34
12.–13. <i>Ján Pavlech</i>	G sv. Jozefa Nové Mesto n. V.	4	2	6	1	–	–	–	–	13	68	34
14. <i>John Richard Ritter</i>	G Masarykovo nám., Třebíč	–	–	–	–	–	–	–	–	–	92	33
15. <i>Jiří Löffelmann</i>	G, Litoměřická, Praha	–	–	–	–	–	–	–	–	–	70	31
16. <i>Andrej Holmes</i>	G Ružomberok	–	–	–	–	–	6	–	–	6	43	24
17. <i>Pavla Trembulaková</i>	G Jírovцова, České Budějovice	1	1	–	–	–	–	–	–	2	42	23
18. <i>Martina Kopecká</i>	G J. Barranda, Beroun	–	–	–	–	–	–	–	–	–	96	22
19. <i>Markéta Jirmanová</i>	BG B. Balbína, Hradec Králové	–	–	–	–	–	–	–	–	–	59	20
20. <i>Filip Keller</i>	G P. de Coubertina, Tábor	–	–	–	–	–	–	–	–	–	72	18
21.–22. <i>Jaroslav Paidar</i>	SPŠ Masarykova, Liberec	–	–	–	–	–	–	–	–	–	52	17
21.–22. <i>Matej Parada</i>	G Jura Hronca, Bratislava	–	–	–	–	–	–	–	–	–	74	17
23. <i>Katarína Žatková</i>	Evanjelické G JAK, Košice	–	–	–	–	–	–	–	–	–	62	16
24. <i>Veronika Funková</i>	G L. Jaroše, Holešov	–	–	–	–	–	–	–	–	–	63	15
25.–26. <i>Dominik Ivan</i>	ŠG PinkHarmony, Zvolen	–	–	–	–	–	–	–	–	–	24	12
25.–26. <i>Ondřej Komora</i>	G Mikulášské n. 23, Plzeň	–	–	–	–	–	–	–	–	–	71	12
27.–28. <i>Kristýna Davídková</i>	Biskupské G, Brno	–	–	–	–	–	–	–	–	–	85	11
27.–28. <i>Jakub Kovářik</i>	G, Hodonín	–	–	–	–	–	–	–	–	–	79	11
29. <i>Ondřej Bílek</i>	SPŠ, Vlašim	–	–	–	–	–	–	–	–	–	100	9
30.–32. <i>Tomáš Hudcovic</i>	Jiráskovo G, Náchod	–	–	–	–	–	–	–	–	–	117	7
30.–32. <i>Štěpán Kastowský</i>	G, Hlučín	–	–	–	–	–	–	–	–	–	64	7
30.–32. <i>Petr Semerák</i>	Jiráskovo G, Náchod	–	–	–	–	–	–	–	–	–	117	7
33.–37. <i>Miroslav Hrabal</i>	G, Olomouc-Hejčín	–	–	–	–	–	–	–	–	–	100	6
33.–37. <i>Štěpán Kohl</i>	Klasické a španělské G, Brno	–	–	–	–	–	–	–	–	–	100	6
33.–37. <i>David Němec</i>	G, Tanvald	–	–	–	–	–	–	–	–	–	100	6
33.–37. <i>Tereza Poláková</i>	G, Budějovická, Praha	–	–	–	–	–	–	–	–	–	27	6
33.–37. <i>Martin Repčák</i>	G, Olomouc-Hejčín	–	–	–	–	–	–	–	–	–	100	6
38. <i>Lukáš Vávah</i>	G J. V. Jirsíka, Č. Budějovice	3	2	0	–	–	–	–	–	5	42	5
39.–41. <i>Ondřej Hajnýš</i>	G, Dvůr Králové n. L.	–	–	–	–	–	–	–	–	–	100	3
39.–41. <i>Jan Lindauer</i>	První české G, Karlovy Vary	–	–	–	–	–	–	–	–	–	100	3
39.–41. <i>Aneta Němcová</i>	G, Boskovice	–	–	–	–	–	–	–	–	–	100	3
42. <i>Filip Geib</i>	G M. M. Hodžu, Liptovský Mikuláš	–	–	–	–	–	–	–	–	–	67	2

Kategorie čtvrtých ročníků

jméno	škola	1	2	3	4	5	P	E	S	V	%	Σ
<i>Student Pilný</i>	MFF UK	3	3	6	8	8	8	12	10	58	<i>100</i>	286
1. Jáchym Bárták	G, Havlíčkův Brod	4	4	7	8	7	6	15	12	63	<i>103</i>	296
2. Daniela Pittnerová	G L. Svobodu, Humenné	4	3	7	9	6	7	9	12	57	<i>94</i>	270
3. Štěpán Stenclík	G, Třinec	-	-	-	-	-	-	-	-	-	<i>102</i>	194
4. Klára Ševčíková	G, Uherské Hradiště	4	3	0	-	5	3	-	3	18	<i>77</i>	172
5. Jan Střeleček	G J. V. Jirsíka, Č. Budějovice	1	3	6	-	-	-	-	-	10	<i>69</i>	146
6. Matěj Mezera	G, Havlíčkův Brod	4	1	-	8	-	-	15	-	28	<i>91</i>	144
7. Filip Novotný	G Masarykovo nám., Kroměříž	3	3	6	1	6	7	-	-	26	<i>77</i>	116
8. Lucie Hronová	G Brno, tř. Kpt. Jaroše	3	3	5	1	5	4	-	-	21	<i>47</i>	111
9. Karel Jókai	G, Špitálská, Praha	-	-	-	-	-	3	-	-	3	<i>59</i>	56
10. Alžběta Andrášková	G, Olomouc-Hejčín	-	-	-	-	-	-	-	-	-	<i>72</i>	48
11. Přemysl Šlastný	G, Žamberk	-	-	-	-	-	-	-	-	-	<i>85</i>	46
12. Jonáš Fuksa	PORG, Praha	-	-	-	-	-	-	-	-	-	<i>82</i>	42
13. Samuel Šipikal	G Milana Ráfusa	-	-	-	-	-	-	-	-	-	<i>53</i>	40
14. Matěj Rzehulka	Wichterlovo G, Ostrava	-	-	-	-	-	-	-	-	-	<i>63</i>	33
15. Petr Šimůnek	G, SOŠ, SOU a VOŠ, Hořice	4	-	-	-	-	-	-	-	4	<i>82</i>	32
16. Martin Vejvoda	G Dobruška	3	3	-	-	-	-	-	-	6	<i>67</i>	31
17. Branislav Belko	G Milana Ráfusa	-	-	-	-	-	-	-	-	-	<i>68</i>	28
18. Veronika Gintnerová	G Sv. Františka, Žilina	-	-	-	-	-	-	-	-	-	<i>68</i>	19
19. František Zach	G, Litomyšl	-	-	-	-	-	-	-	-	-	<i>100</i>	18
20. Kateřina Stodolová	G Dašická, Pardubice	-	-	-	-	-	-	-	-	-	<i>87</i>	13
21. Petra Štefaníková	G O. Havlové, Ostrava	-	-	-	-	-	-	-	-	-	<i>91</i>	10
22. Václav Mikeska	G F. Palackého, Val. Mez.	-	-	-	-	-	-	-	-	-	<i>100</i>	6
23. Tomáš Tesař	G Jana Keplera, Praha	-	-	-	-	-	-	-	-	-	<i>27</i>	4


**FYKOS**

UK, Matematicko-fyzikální fakulta

Ústav teoretické fyziky

V Holešovičkách 2

180 00 Praha 8

www: <http://fykos.cz>e-mail: fykos@fykos.czFYKOS je také na Facebooku <http://www.facebook.com/Fykos>

Fyzikální korespondenční seminář je organizován studenty MFF UK. Je zastřešen Oddělením pro vnější vztahy a propagaci MFF UK a podporován Ústavem teoretické fyziky MFF UK, jeho zaměstnanci a Jednotou českých matematiků a fyziků.

Toto dílo je šířeno pod licencí Creative Commons Attribution-Share Alike 3.0 Unported. Pro zobrazení kopie této licence navštivte <http://creativecommons.org/licenses/by-sa/3.0/>.