Úloha IV.S ... testovací

10 bodů; průměr 7,56; řešilo 25 studentů

a) *Zkuste vlastními slovy popsat, k čemu a jak se používá testování hypotéz (postačí vlastními slovy popsat následující: hypotéza a alternativa, chyba 1. a 2. druhu, hladina testu, testová statistika, kritický obor testu, p-hodnota testu pro konkrétní naměřená data). Není potřeba uvádět přesná matematická odvození, stačí požadované pojmy a vlastnosti stručně popsat.*

b) *V přiloženém datovém souboru* testovani1.csv *najdete naměřené hodnoty určité fyzikální veličiny. Pomocí jednovýběrového t-testu otestujte, zda je skutečná hodnota měřené fyzikální veličiny rovna* 20*. Dále předpokládejme, že je naším cílem ukázat, že hodnota měřené fyzikální veličiny je větší než* 20*. Použijte vhodnou jednostrannou modifikaci t-testu k tomu, abyste toto tvrzení ověřili (dejte si pozor na správné zvolení hypotézy a alternativy).*

c) *V přiloženém datovém souboru* testovani2.csv *najdete naměřené hodnoty 2 různých fyzikálních veličin. Představujme si, že se jedná o měření stejné fyzikální charakteristiky ale za různých vnějších podmínek (teplota, tlak atd.). Pomocí dvouvýběrového z-testu otestujte hypotézu, že hodnota této fyzikální charakteristiky je pro obě volby vnějších podmínek stejná.*

d) *Použijte stejná data jako v seriálové úloze z první série a pomocí Kolmogorovova-Smirnovova testu určete, který ze 4 vzorků dat pochází z normálního rozdělení a který vzorek pochází z exponenciálního rozdělení.*

Bonus *Předpokládejte, že máte k dispozici měření 2 fyzikálních veličin (tedy 2 sady naměřených hodnot), kde jsou všechna měření na sobě nezávislá. Odvoďte upravený dvouvýběrový z-test, který by testoval hypotézu, že skutečná hodnota první měřené fyzikální veličiny je dvojnásobek skutečné hodnoty druhé měřené fyzikální veličiny. Pro udělení bodů je nutné a postačuje odvodit podobu testové statistiky a kritického oboru. (*Nápověda: *Použijte vícerozměrnou verzi CLV, kde vhodně zvolíte funkci f, a dále postupujte analogicky jako u odvození klasického dvouvýběrového z-testu)*

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 chtěl otestovat, jak těžké úlohy řešitelé zvládnou.

a) Detailní odpověď na tuto otázku dostanete pouze přečtením 3. dílu seriálu, v tomto vzorovém řešení uvedeme jen ty nejdůležitější věci.

Testování hypotéz se použije v případě, kdy potřebujeme na základě naměřených dat rozhodnout o platnosti nějaké naší domněnky (hypotézy) nebo tvrzení. Jak už jsme rozebírali v předchozích dílech seriálu, reprezentujeme naměřená data jako realizace určité náhodné veličiny (tedy naměřená data jsou náhodná - při každém měření dostáváme jiné hodnoty i když používáme naprosto stejný postup měření). Testování hypotéz lze chápat jako postup, který nám říká, zda jsou naměřená data dostatečně reprezentativní a zda na jejich základě můžeme pronášet závěry o platnosti testované hypotézy (případně jaké jsou pravděpodobnosti, že se dopouštíme špatného závěru).

Hypotéza a alternativa nám říkají, která tvrzení budeme testovat. Vždy je nutné k testované hypotéze sestavit alternativu, která bude opačné tvrzení než hypotéza (tj. platí buď hypotéza nebo alternativa, nikdy obě současně a vždy alespoň jedna z nich). Hypotézu a alternativu musíme vždy zapsat pomocí matematických výrazů (nejčastěji rovnosti nebo nerovnosti) a parametrů uvažovaného modelu.

Jelikož jsou měřená data náhodná, můžeme (ač s malou pravděpodobností) naměřit v podstatě jakákoliv data a nikdy si nemůžeme být úplně jisti, jestli je závěrečné rozhodnutí správné. Při formulaci závěru můžeme udělat celkem 2 druhy chyb (označili jsme si je jako chyba 1. a 2. druhu). Chyba 1. druhu znamená, že jsme zamítli hypotézu, která ale ve skutečnosti platí. Chyba 2. druhu znamená, že jsme nezamítli hypotézu, která ve skutečnosti neplatí (viz Tab. 1).

naše rozhodnutí / skutečný stav	platí hypotéza	platí alternativa
zamítáme hypotézu	chyba 1. druhu	správně
nezamítáme hypotézu	správně	chyba 2. druhu

Tab. 1: Tabulka možných rozhodnutí a chyb

Hladina testu udává pravděpodobnost chyby 1. druhu, značí se obvykle α a její hodnota se obvykle volí jako $\alpha = 0{,}05$ (v případě důležitých experimentů nižší). Tuto hodnotu si musíme vždy zvolit na začátku zpracování dat.
Testová statistika je transformace naměřených dat, na základě které se budeme následně rozhodovat o (ne)zamítání hypotézy. Tato transformace musí být zvolena tak, abychom znali její rozdělení za platnosti testované hypotézy (za předpokladu, že máme dostatečný počet měření, stačí asymptotické rozdělení).
Kritický obor testu je množina hodnot testové statistiky (většinou interval nebo sjednocení dvou intervalů), ve kterém za platnosti hypotézy leží hodnota testové statistiky jen s pravděpodobností α (hladina testu). Jinými slovy je to množina, kde by za platnosti hypotézy hodnota testové statistiky nejspíše ležet neměla, ale za platnosti alternativy je velká pravděpodobnost, že zde bude hodnota testové statistiky ležet. Když máme v praxi provést nějaký test s použitím naměřených dat, stačí spočítat realizovanou hodnotu testové statistiky a zjistit, zda leží v kritickém oboru nebo nikoliv. Pokud leží realizovaná hodnota testové statistiky v kritickém oboru potom zamítáme platnost hypotézy, v opačném případě platnost hypotézy nezamítáme.
p-hodnota testu pro konkrétní naměřená data představuje jakýsi alternativní způsob rozhodování o (ne)zamítání platnosti hypotézy (výsledek je vždy naprosto stejný jako v případě postupu na základě kritického oboru, ale postup rozhodování je jiný). Když máme konkrétní naměřená data, potom definujeme p-hodnotu jako pravděpodobnost, že bychom za platnosti hypotézy naměřili data, která by ještě více protiřečila testované hypotéze než ta data, která jsme naměřili. Rozhodování o (ne)zamítání platnosti hypotézy na základě p-hodnoty je jednoduché, hypotézu zamítáme právě tehdy když je p-hodnota menší než námi předem určená hladina testu α (v tomto případě je pravděpodobnost naměření dat ještě více protiřečících hypotéze malé, tedy jsme naměřili data, která hypotéze hodně protiřečí). Statistické programy obvykle upřednostňují rozhodování na základě p-hodnoty, kterou většinou spočítají numericky.

b) Skutečnou hodnotu měřené fyzikální veličiny označíme jako μ_x. Jelikož považujeme naměřená data za realizace náhodné veličiny s normálním rozdělením, střední hodnotou μ_x a konečným rozptylem můžeme na testování takovéto hypotézy použít jednovýběrový t-test. Zvolená hladina bude standardně rovna $\alpha = 0{,}05$. Testová hypotéza a alternativa bude ve tvaru

$$H : \mu_x = 20\,,$$
$$A : \mu_x \neq 20\,.$$

Realizovaná hodnota testové statistiky při použití přiložených dat vyjde jako

$$T_n = \sqrt{n}\frac{\overline{x_n} - \mu_0}{S_n} \doteq 2{,}52\,.$$

Kritický obor testu je tvaru

$$C = (-\infty, t_{n-1,\frac{\alpha}{2}}) \cup (t_{n-1,1-\frac{\alpha}{2}}, \infty)\,,$$
$$C = (-\infty, -2{,}06) \cup (2{,}06, \infty)$$

a příslušná p-hodnota pro naměřená data vyšla

$$p \doteq 0{,}0187\,.$$

Na základě porovnání realizované hodnoty testové statistiky s kritickým oborem a na základě p-hodnoty dospíváme k závěru, že zamítáme platnost hypotézy. Jako závěr tedy uvedeme, že na základě naměřených dat můžeme prohlásit (na hladině spolehlivosti $\alpha = 0{,}05$), že hodnota měřené fyzikální veličiny není rovna $\mu_0 = 20$.

Při řešení druhé části úkolu je důležité správně zvolit testovou hypotézu a alternativu. Volbu budeme provádět přesně tak, jak je popsáno v seriálu a sice tak, že budeme chtít vyvrátit opak tvrzení, které chceme potvrdit. Budeme tedy usilovat o to vyvrátit platnosti tvrzení, že skutečná hodnota měřené fyzikální je menší než 20. Toto opačné tvrzení, které chceme vyvrátit tedy zvolíme jako testovanou hypotézu. Testová hypotéza a alternativa tedy budou ve tvaru

$$H : \mu_x \leq 20\,,$$
$$A : \mu_x > 20\,.$$

Realizovaná hodnota testové statistiky při použití přiložených dat vyjde stejně jako v předchozím případě, neboť používáme stejnou testovou statistiku

$$T_n = \sqrt{n}\frac{\overline{x_n} - \mu_0}{S_n} \doteq 2{,}52\,.$$

Kritický obor tohoto testu se ale bude lišit

$$C = (t_{n-1,1-\alpha}, \infty)\,,$$
$$C = (1{,}71, \infty)$$

a příslušná p-hodnota pro naměřená data vyšla

$$p \doteq 0{,}009\,.$$

Vidíme tedy, že rozhodnutí (jak na základě porovnání realizované hodnoty testové statistiky s kritickým oborem tak na základě p-hodnoty) je zamítnout platnost testované hypotézy. Tímto jsme tedy potvrdili platnost alternativy (na hladině spolehlivosti $\alpha = 0{,}05$), tedy tvrzení že skutečná hodnota měřené fyzikální veličiny je větší než 20.

c) Označíme μ_X a μ_Y skutečné hodnoty měřené fyzikální charakteristiky za dvou různých vnějších podmínek. Cílem našeho testu bude otestovat hypotézu, že jsou tyto dvě hodnoty stejné, k čemuž použijeme dvouvýběrový z-test. Tento test byl podrobně popsán v textu seriálů, proto ho už na tomto místě nebudeme podrobně popisovat a jen popíšeme, jak se na takovýto případ konkrétně použije. Testovaná hypotéza a alternativa mají následující podobu

$$\begin{aligned} H &: \mu_X - \mu_Y = 0\,, \\ A &: \mu_X - \mu_Y \neq 0\,. \end{aligned}$$

Realizovaná hodnota testové statistiky za použití naměřených dat vychází jako

$$Z_{n,m} = \frac{\overline{x_n} - \overline{y_m} - \vartheta}{\sqrt{\frac{S^2_{n_x}}{n} + \frac{S^2_{m_y}}{m}}} \doteq 3{,}02\,.$$

Kritický obor je v tomto případě ve tvaru

$$\begin{aligned} C &= (-\infty, u_{\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}, \infty)\,, \\ C &= (-\infty, -1{,}96) \cup (1{,}96, \infty). \end{aligned}$$

p-hodnota testu pro naše naměřená data nabývá hodnoty

$$p \doteq 0{,}0043\,.$$

Vidíme, že na základě porovnání realizované hodnoty testové statistiky s kritickým oborem i na základě p-hodnoty testu zamítáme testovanou hypotézu na hladině spolehlivosti $\alpha = 0{,}05$. Závěrem tedy lze říci, že jsme prokázali (na hladině spolehlivosti $\alpha = 0{,}05$), že hodnota měřené fyzikální charakteristiky závisí na vnějších podmínkách (tedy za dvou různých zkoumaných vnějších podmínek je hodnota této fyzikální charakteristiky navzájem odlišná).

d) Budeme používat jednovýběrový Kolmogorovův-Smirnovův test, který byl popsán v textu seriálu a ve vzorovém skriptu, zde jen připomeneme jeho nejdůležitější charakteristiky. Nechť máme nějaké spojité rozdělení S (v našem případě normální a exponenciální rozdělení) a naměřená data, potom můžeme pomocí Kolmogorovova-Smirnovova testu otestovat následující hypotézu a alternativu

$$\begin{aligned} H &: \text{měřená data mají rozdělení } S\,, \\ A &: \text{měřená data nemají rozdělení } S\,. \end{aligned}$$

Testovou statistiku a kritický obor testu už v tomto případě uvádět nebudeme, neboť jsou příliš komplikované[1], postačí nám výstup ze statistického programu R, kde bude uvedena p-hodnota pro konkrétní naměřená data.

Postupovat budeme tak, že pro každou sadu naměřených hodnot (celkem 4 sady měření) provedeme dvakrát Kolmogorovův-Smirnovův test, kde budeme rozdělení S volit jako normální a exponenciální. Pokud pro konkrétní sadu naměřených hodnot Kolmogorovův-Smirnovův

[1] Hloubaví čtenáři si mohou tyto údaje vyhledat na internetu, např. `https://en.wikipedia.org/wiki/Kolmogorov-Smirnov_test`.

test zamítne hypotézu, že tato data pocházejí z normálního (nebo exponenciálního) rozdělení, můžeme si být poměrně jistí, že z tohoto rozdělení opravdu nepocházejí (na hladině $\alpha = 0{,}05$). V opačném případě (tedy nezamítnutí hypotézy) sice nemůžeme naše výsledky interpretovat jako přímé potvrzení testované hypotézy, ale budeme v tomto případě považovat za dostatečný důvod k tomu tvrdit, že hypotéza je platná[2]. Výsledky takovýchto testů můžeme vidět v Tabulce 2.

Na základě těchto výsledků vidíme, že jediný vzorek, u kterého jsme nezamítali hypotézu o normálním rozdělení, je Vzorek 3. Podobně jediný vzorek, u kterého jsme nezamítali hypotézu o exponenciálním rozdělení, je Vzorek 4. Jak už bylo diskutováno dříve, budeme tyto výsledky považovat za dostatečné k tomu, abychom tvrdili, že data ve Vzorku 3 pocházejí z normálního rozdělení a data ve Vzorku 4 pocházejí z exponenciálního rozdělení. Pozorný řešitel tyto výsledky porovná se závěry o rozdělení, ze kterého měřená data pocházejí, které jsme udělali na základě histogramů v seriálové úloze v 1. sérii, a zjistí, že jsou tyto závěry stejné.

Tab. 2: *p*-hodnoty Komogorovova-Smirnonovova testu pro všechny sady měřených dat a testovaných rozdělení.

	Normální rozdělení	Exponenciální rozdělení
Vzorek 1	$< 2{,}2 \cdot 10^{-16}$	$< 2{,}2 \cdot 10^{-16}$
Vzorek 2	$< 2{,}2 \cdot 10^{-16}$	$< 2{,}2 \cdot 10^{-16}$
Vzorek 3	0,92	$< 2{,}2 \cdot 10^{-16}$
Vzorek 4	$< 2{,}2 \cdot 10^{-16}$	0,46

Bonus: Na tomto místě si musíme připomenout znění vícerozměrné centrální věty, která byla uvedena ve 3. díle seriálu. Vícerozměrná verze CLV říká, že pokud měříme několik fyzikálních veličin, čímž dostaneme odhady těchto veličin s nejistotou měření

$$\left(\overline{v^{(1)}} \pm s_{n_1}^{(1)}\right), \ldots, \left(\overline{v^{(k)}} \pm s_{n_k}^{(k)}\right),$$

potom pro tyto naměřené hodnoty a libovolnou diferencovatelnou funkce f bude platit

$$\frac{f\left(\overline{v_n^{(1)}}, \ldots, \overline{v_n^{(k)}}\right) - f\left(v^{(1)}, \ldots, v^{(k)}\right)}{\sqrt{S^2}} \xrightarrow{D} N(0,1)\,,$$

kde $v^{(1)}, \ldots, v^{(k)}$ jsou skutečné hodnoty měřených fyzikálních veličin a výraz S^2 ve jmenovateli je určen vzorcem

$$S^2 = \left(\frac{\partial f}{\partial v^{(1)}}(\overline{v}), \quad \cdots \quad, \frac{\partial f}{\partial v^{(k)}}(\overline{v})\right) \begin{pmatrix} s_{n_1}^2 & \cdots & \frac{\widehat{\mathrm{cov}}\left(v^{(1)}, v^{(k)}\right)}{\sqrt{n_1 n_k}} \\ \frac{\widehat{\mathrm{cov}}\left(v^{(2)}, v^{(1)}\right)}{\sqrt{n_2 n_1}} & \cdots & \frac{\widehat{\mathrm{cov}}\left(v^{(2)}, v^{(k)}\right)}{\sqrt{n_2 n_k}} \\ \vdots & \ddots & \vdots \\ \frac{\widehat{\mathrm{cov}}\left(v^{(k)}, v^{(1)}\right)}{\sqrt{n_k n_1}} & \cdots & s_{n_k}^2 \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial v^{(1)}}(\overline{v}) \\ \vdots \\ \frac{\partial f}{\partial v^{(k)}}(\overline{v}) \end{pmatrix}.$$

[2]V praxi se výsledky takovýchto testů ještě kombinují s dalšími metodami (např. zkoumáním pomocí histogramů) a na základě kombinovaných výsledků se formuluje závěr.

Jelikož pracujeme se dvěma fyzikálními veličinami, budeme používat dvourozměrnou verzi CLV (tedy $k = 2$). Označíme μ_X a μ_Y skutečné hodnoty měřených fyzikálních veličin, naším cílem bude otestovat, zda platí

$$\mu_X = 2\mu_Y .$$

Jako alternativu budeme uvažovat jednoduše negaci této hypotézy. Hypotézu a alternativu můžeme ekvivalentně zapsat jako

$$\begin{aligned} H :& \mu_X - 2\mu_Y = 0\,, \\ A :& \mu_X - 2\mu_Y \neq 0\,. \end{aligned}$$

Nyní budeme předpokládat, že naše testová hypotéza platí a ve dvourozměrné CLV zvolíme funkci f ve tvaru

$$f(x,y) = x - 2y.$$

Tuto volbu děláme proto, aby nám za platnosti testované hypotézy vyšlo

$$f(\mu_X, \mu_Y) = 0.$$

Testová statistika, kterou označíme jako T, bude úplně stejná jako výraz ve znění CLV, tedy

$$T = \frac{f\left(\overline{\mu_X}, \overline{\mu_Y}\right) - f(\mu_X, \mu_Y)}{\sqrt{S^2}} = \frac{\overline{\mu_X} - 2\overline{\mu_Y}}{\sqrt{S^2}}, \tag{1}$$

kde $\overline{\mu_X}, \overline{\mu_Y}$ označují výběrové průměry měření první a druhé fyzikální veličiny. Pro tuto speciální volbu funkce f také dostaneme

$$S^2 = \begin{pmatrix} 1 & \text{-}2 \end{pmatrix} \begin{pmatrix} s^2_{n_1} & \frac{\widehat{\mathrm{cov}}(X,Y)}{\sqrt{n_1 n_2}} \\ \frac{\widehat{\mathrm{cov}}(X,Y)}{\sqrt{n_1 n_2}} & s^2_{n_2} \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} .$$

Z vícerozměrné CLV plyne, že testová statistika (1) za platnosti testované hypotézy konverguje v distribuci k rozdělení $N(0,1)$. Určení podoby kritického oboru testu pro hladinu testu α bude tedy velmi podobné jako odvození podoby kritického oboru testu pro klasický dvouvýběrový z-test (podrobně popsáno v textu seriálu). Kritický obor tedy bude mít tvar

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

po vyčíslení pro $\alpha = 0{,}05$

$$C = (-\infty, -1{,}96) \cup (1{,}96, \infty)\,. \tag{2}$$

Pozorný řešitel si sám rozmyslí, že pravděpodobnost, že za platnosti alternativy padne realizovaná hodnota testové statistiky do kritického oboru se se zvyšujícím počtem měření blíží 1 (naprosto stejná úvaha jako je použita v seriálu při odvozování podobu kritického oboru pro dvouvýběrový z-test). Kritický obor je tedy zvolen optimálně.

Závěrem tedy můžeme říci, že jsme odvodili podobu statistického testu k testování hypotézy, že jedna měřená fyzikální veličina má dvojnásobnou hodnotu než druhá měřená fyzikální veličina, testová statistika tohoto testu je tvaru (1) a kritický obor je tvaru (2). Na závěr poznamenejme, že úplně analogicky by se dal odvodit podobný test, kde bychom testovali hypotézu, že hodnota jedné měřené fyzikální veličiny je k násobkem hodnoty druhé měřené fyzikální veličiny (kde k může být libovolné reálné číslo).

Michal Nožička
nozicka@fykos.cz