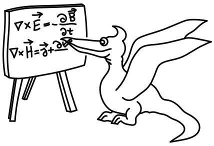

Seriál: Zpracování dat fyzikálních měření

V tomto díle seriálu se budeme věnovat statistickému testování hypotéz. Tento díl bude výrazným způsobem navazovat na všechny 3 předchozí díly seriálu, proto doporučujeme si nejprve zběžně připomenout obsah minulých dílů. Nyní už k samotnému tématu tohoto dílu seriálu: co to je statistické testování hypotéz?

Jistě se každý mnohokrát setkal se situací, kdy si potřeboval na základě naměřených dat ověřit nějakou domněnku, hypotézu. Jako fyzikální příklad můžeme uvést třeba porovnávání rychlostí zvuku v různých prostředích (chceme například testovat domněnku, že se zvuk v železe šíří rychleji než ve vzduchu) nebo můžeme chtít zjistit, jestli má například namokření povrchu vliv na koeficient tření a podobně.

Naše závěry musí být vždy založeny na naměřených datech. Kdybychom uměli fyzikální veličiny měřit přesně, tak by odpověď na položenou otázku byla velice jednoduchá, stačilo by příslušné fyzikální veličiny změřit a rozhodnout o platnosti hypotézy. V praxi to ovšem tak jednoduše udělat nemůžeme, neboť nikdy nebudeme mít naprosto přesná měření. Jak si jistě z minulých dílů pamatujete, naměřená data v našem matematickém modelu považujeme za realizace náhodné veličiny. Problém tedy spočívá v tom, že teoreticky můžeme naměřit jakákoliv data (i když třeba s velice malou pravděpodobností). Následně je potřeba se rozhodovat na základě takovýchto „náhodných“ dat, což není jednoduchý úkol. Představme si například, že měříme dvě fyzikální veličiny a následně je chceme porovnat (tj. zjistit, která je větší). Naměřená data mohou vypadat například jako data v tabulce 1.

Na první pohled je vidět, že data naměřená ve 2. vzorku mají mírně vyšší hodnoty (alespoň tedy v průměru). Jak ale poznat, že je to opravdu způsobeno vyšší hodnotou měřené fyzikální veličiny a ne pouze tím, že jsme zrovna náhodně naměřili takováto data (vzpomeňte si na to, že naměřenou hodnotu považujeme za náhodnou veličinu, proto při každém měření můžeme naměřit jinou hodnotu)? Tomuto procesu se říká testování hypotéz.

Tab. 1: Hypotetická naměřená data.

1. vzorek	8	7	9	8	8	6	10	7	8	7
2. vzorek	9	8	9	7	10	8	8	7	9	7

Základní idea testování hypotéz

Nyní už si budeme postup statistického testování hypotéz popisovat matematicky přesněji, než jsme to udělali v úvodu, kde jsme jenom nastínili obecný problém.

Naše otázka bude vždy formulovaná tak, že budeme mít nějaký výrok, který budeme označovat za hypotézu (například výrok „Namokření podložky nemá žádný vliv na třecí koeficient.“), kterou budeme zkráceně značit H. K hypotéze si musíme zkonstruovat opačný výrok, který budeme označovat jako alternativu (v našem případě by to byl výrok „Namokření podložky má vliv na třecí koeficient.“), kterou budeme označovat jako A. Je důležité dát si pozor na to, aby alternativa byla opravdu opačný výrok k hypotéze [1]. V našem případě by například výrok „Namokření podložky snižuje třecí koeficient“ nemohl být považován za alternativu (může totiž platit ještě něco jiného než hypotéza a alternativa, tj. výrok „Namokření podložky zvyšuje třecí koeficient.“).

Testování hypotéz bude vlastně proces, jak budeme na základě naměřených dat (tedy realizací náhodných veličin) rozhodovat o platnosti hypotézy nebo alternativy. V případě, že budeme mít vyhovujícím způsobem zformulovanou hypotézu a alternativu, budeme postupovat tak, že budeme zamítat nebo nezamítat [2] platnost hypotézy. Musíme se smířit s tím, že nikdy nebudeme schopni s určitostí říci, jestli hypotéza platí nebo ne. Je to způsobeno tím, že považujeme výsledky měření za náhodnou veličinu, a proto můžeme naměřit (i když s velice malou pravděpodobností) v podstatě jakákoliv data. Následně nemáme žádné prostředky na to, jak posoudit, zda tato data jsou skutečně reprezentativní, nebo zda jde jen o velkou anomálii.

Když se budeme rozhodovat, jak při rozhodování o platnosti hypotézy postupovat, musíme si nejprve uvědomit, jaké všechny možnosti mohou nastat a zanalyzovat si, jaké chyby můžeme případně udělat. Na základě této analýzy budeme potom navrhovat postup testování hypotéz.

Jak je z tabulky 2 vidět, můžeme při našem rozhodování udělat 2 druhy chyb, které jsme si označili jako chyba 1. a 2. druhu. Nyní se zaměříme na to, jak těmto chybám čelit a zvolíme takový postup, kdy se budeme s největší pravděpodobností rozhodovat správně.

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

Tab. 2: Tabulka možných rozhodnutí a chyb

Základní ideou testování hypotéz je zvolit takový postup rozhodování o platnosti hypotézy, aby byla pravděpodobnost chyby 1. druhu rovna nějaké předem zvolené konstantě α (většinou se v praxi volí hodnota $\alpha = 0{,}05$) a pravděpodobnost chyby 2. druhu byla co možná nejmenší. Na první pohled se takovýto postup může zdát komplikovaný a neoptimální, ale je to opravdu to nejlepší, co můžeme udělat [3].

Základní postup při odvozování testů

Nyní už podrobněji k samotnému postupu rozhodování o (ne)zamítnutí hypotézy. Nejprve zde poskytneme základní postup odvozování statistických testů a následně si odvodíme nejčastěji používané testy.

První věc, kterou musíme při odvozování statistických testů udělat, je vhodně zapsat testovanou hypotézu a alternativu. Vhodným zápisem se rozumí zápis pomocí rovností nebo nerovností určitých parametrů našeho modelu.

Následně musíme zvolit testovou statistiku. Testová statistika bude určitá transformace našich naměřených dat (tím pádem to bude náhodná veličina, neboť naše měřená data považujeme za náhodné veličiny), o které budeme vědět, jak se chová za platnosti hypotézy i alternativy.

[1] Tj. aby se nemohlo stát, že bude ve skutečnosti platit ještě něco úplně jiného, ani že budou platit hypotéza i alternativa zároveň.

[2] Nezamítnout neznamená to samé co potvrdit platnost!

[3] Zejména je nutné si uvědomit, že opravdu nemáme žádný nástroj na přímé ověření, zda hypotéza platí. Jediné, co můžeme dělat, je otestovat tuto hypotézu oproti vhodné alternativě.

Tím se zejména myslí, že budeme v obou případech znát její rozdělení. Testová statistika je ve většině případů pouze jednorozměrná náhodná veličina, tedy použitá transformace z n naměřených dat určitým způsobem (určitým vzorečkem) vyrobí pouze jedno číslo (realizaci testové statistiky).

Posledním krokem bude volba tzv. kritického oboru testu. Když víme, jaké rozdělení by měla mít testová statistika za platnosti hypotézy a za platnosti alternativy, můžeme se rozhodnout, zda realizovaná hodnota testové statistiky odpovídá spíše hypotéze či alternativě. To provedeme tak, že si pro předem zvolenou hladinu testu α určíme interval, ve kterém bude za platnosti hypotézy s pravděpodobností $1-\alpha$ ležet realizovaná hodnota testové statistiky. Takovýchto intervalů samozřejmě bude existovat nekonečně mnoho, my se ale budeme zajímat o takový interval, aby za platnosti alternativy ležela realizovaná hodnota testové statistiky v tomto intervalu s co nejmenší pravděpodobností. Doplněk tohoto intervalu budeme označovat jako kritický obor testu. Konečné rozhodnutí je potom jednoduché, pokud leží realizovaná hodnota testové statistiky v kritickém oboru, zamítneme hypotézu ve prospěch alternativy, pokud realizovaná hodnota testové statistiky neleží v kritickém oboru, nebudeme zamítat testovanou hypotézu.

Každý statistický test je tedy určen čtveřicí hypotéza, alternativa, testová statistika a kritický obor.

Je nutné si uvědomit, že testování hypotéz nedává jednoznačné odpovědi a kvůli náhodnosti měřených dat ani dávat nemůže. Vždy budeme pracovat na určité hladině spolehlivosti α (obvykle se volí hladina $\alpha = 0{,}05$) a všechny naše závěry tedy bude nutné interpretovat tak, že je pravděpodobnost α, že jsme udělali chybu 1. druhu. O volbě hladiny spolehlivosti testů v konkrétních případech budeme mluvit ještě v další části tohoto textu, nyní už si můžeme odvodit několik nejzákladnějších statistických testů.

Jednovýběrový t-test

Při jednovýběrovém t-testu předpokládáme, že měřená data mají normální rozdělení (s libovolnými parametry) a že jsou jednotlivá měření na sobě nezávislá. Jednovýběrový t-test testuje hypotézu, že měřená fyzikální veličina (tedy střední hodnota měřených dat) je rovna nějaké předem zvolené konstantě, oproti alternativně, že střední hodnota měřené náhodné veličiny této konstantě rovna není. Zkráceně se často píše

$$H : \mu_x = \mu_0\,,$$
$$A : \mu_x \neq \mu_0\,,$$

kde μ_x označuje skutečnou hodnotu střední hodnoty měřené náhodné veličiny (tuto hodnotu v praxi téměř nikdy neznáme), μ_0 označuje hypotetickou hodnotu střední hodnoty měřené náhodné veličiny (tuto hodnotu dopředu známe) a písmena H, A označují hypotézu a alternativu.

Nyní si můžeme podrobně popsat odvození testové statistiky jednovýběrového t-testu. Vyjdeme z předpokladu, že za platnosti hypotézy bude mít transformace naměřených dat

$$T_n = \sqrt{n}\frac{\overline{x_n} - \mu_0}{S_n}\,, \tag{1}$$

kde $\overline{x_n}$, resp. S_n, označují výběrový průměr, resp. výběrovou směrodatnou odchylku, Studentovo rozdělení o $n-1$ stupních volnosti t_{n-1}. Tento fakt byl podrobně popsán v druhém dílu seriálu.

Nyní si stačí uvědomit, že za platnosti hypotézy bude pravděpodobnost, že testová statistika T_n nabyde hodnoty z intervalu

$$(t_{n-1,\frac{\alpha}{2}}, t_{n-1,1-\frac{\alpha}{2}})\,, \tag{2}$$

rovna přesně $1-\alpha$. Toto tvrzení vychází přímo z definice kvantilu Studentova rozdělení (podobný koncept už se používal v minulých dílech seriálu). Zároveň bude platit, že za platnosti alternativy bude s největší pravděpodobností hodnota testové statistiky výrazně odlišná od nuly. Toto je způsobeno tím, že $\overline{x_n}$ bude pro velká n s největší pravděpodobností velice blízká skutečné střední hodnotě μ_x a tím pádem bude čitatel zlomku (1) s největší pravděpodobností velice blízký hodnotě

$$\Delta = \mu_x - \mu_0\,.$$

Jmenovatel zlomku (1) bude pro velká n s největší pravděpodobností velice blízký skutečné směrodatné odchylce. Tedy hodnota celého zlomku (1) bude pro velká n s největší pravděpodobností velice blízká hodnotě

$$\frac{\Delta}{\sigma} \neq 0\,.$$

Po přenásobení této hodnoty číslem $\sqrt{n}$ potom bude hodnota testové statistiky T_n buď hodně velká, nebo hodně malá (hodně velké záporné číslo), každopádně bude s největší pravděpodobností výrazně odlišné od nuly. Tento poznatek nás opravňuje ke stanovení kritického oboru C jako doplňku intervalu (2), tedy jako sjednocení intervalů

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

Na závěr k tomuto testu jen uveďme, že pokud nebude splněna podmínka na normální rozdělení měřených dat, bude tento test fungovat asymptoticky (tj. pro velký počet měření, cca víc než 30). Toto vychází z platnosti centrální limitní věty (viz minulý díl seriálu), díky níž bude mít testová statistika T_n pro velký počet měření přibližně rozdělení $N(0,1)$ a kvantily Studentova t_{n-1} rozdělení a rozdělení $N(0,1)$ budou pro takto velké n prakticky shodné.

Jen drobnou úpravou by se dal tento test upravit na testování dvojice hypotéza a alternativa v následujícím tvaru

$$H : \mu_x \geq \mu_0\,,$$
$$A : \mu_x < \mu_0\,.$$

V tomto případě bychom použili stejnou testovou statistiku T_n, ale volba kritického oboru by se mírně lišila. Podobným způsobem jako v případě klasického t-testu bude za předpokladu, že platí $\mu_x > \mu_0$ (tj. za platnosti hypotézy[4]), hodnota testové statistiky pro velká n velmi velká (neboť hodnota zlomku bude s největší pravděpodobností kladná a po přenásobení členem $\sqrt{n}$ bude hodnota testové statistiky velmi velká). Naopak za platnosti alternativy (tj. v případě, že platí $\mu_x < \mu_0$) bude hodnota testové statistiky s největší pravděpodobností velké záporné číslo, což lze odvodit podobnou úvahou. Je tedy jasné, že pro velké hodnoty realizace testové statistiky hypotézu zamítat nebudeme a naopak pro velmi malé hodnoty (tj. velké záporné hodnoty) testové statistiky budeme zamítat hypotézu. Co jsou to ale ty „velmi malé hodnoty“? To lze odvodit na základě toho, jak se testová statistika (resp. její rozdělení) chová v hraničním případě, kdy platí $\mu_x = \mu_0$. Jak už bylo dříve odvozeno, testová statistika má v tomto případě Studentovo rozdělení o $n-1$ stupních volnosti t_{n-1}. Kritický obor tohoto testu tedy musíme

[4] Ve skutečnosti se do platnosti hypotézy ještě vejde případ, kdy $\mu_x = \mu_0$, ale to budeme řešit ještě později.

zvolit tak, aby v tomto hraničním případě byla pravděpodobnost, že realizovaná hodnota testové statistiky bude ležet v kritickém oboru, rovna hladině testu α. Kritický obor tedy bude mít tvar

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

čímž bude tato podmínka splněna a zároveň bude platit, že za platnosti alternativy pro velká n bude realizovaná hodnota testové statistiky s největší pravděpodobností ležet v kritickém oboru.

Opět platí, že tento test lze použít v případě, kdy měřená data nemají normální rozdělení, ale musíme se smířit s tím, že bude fungovat pouze asymptoticky. Je vhodné ho použít jen v případě, kdy máme k dispozici alespoň 30 měření, jinak mohou jeho výsledky být nepřesné.

Analogickým způsobem by se dal tento test upravit pro testování hypotézy s opačnými znaménky nerovností (použili bychom stejnou testovou statistiku a hypotézu bychom zamítali pouze pro hodnoty realizované testové statistiky větší než $t_{n-1,1-\alpha}$).

Na závěr jen uvedeme, že v praxi můžeme potkat tzv. pravidlo $n\sigma$ (za n dosaďme čísla $1, 2, 3, \ldots$). Toto pravidlo je vlastně jakýmsi ekvivalentním vyjádřením toho, co dělá t-test, a sice říká, že bychom měli hypotézu zamítat, pokud rozdíl hypotetické hodnoty a hodnoty odhadnuté z naměřených dat (tj. výběrového průměru) bude větší než s_n, případně $2s_n, 3s_n \ldots$. Rozmyslete si sami, že rozhodování na základě tohoto pravidla je ekvivalentní t-testu pro speciální volby $\alpha = 0{,}32$, případně $\alpha = 0{,}05$ atd.

Dvouvýběrový z-test

Dvouvýběrový z-test se použije v případech, kdy chceme porovnat hodnotu dvou měřených fyzikálních veličin [5]. Předpokládáme, že naměřená data budou mít tedy následující podobu: $x_1, \ldots, x_n$ je měření první veličiny a $y_1, \ldots, y_m$ je měření druhé veličiny. Předpokládáme, že jsou všechna měření na sobě nezávislá a neděláme žádné předpoklady o rozdělení měřených dat[6].

Na základě těchto měření budeme chtít porovnat skutečné hodnoty těchto dvou fyzikálních veličin (které se rovnají středním hodnotám dvou měřených náhodných veličin, jak bylo ukázáno v předchozích dílech seriálu). Budeme chtít testovat hypotézu

$$\begin{aligned} H :& \mu_x - \mu_y = \vartheta\,, \\ A :& \mu_x - \mu_y \neq \vartheta\,, \end{aligned}$$

kde μ_x je skutečná střední hodnota 1. měřené náhodné veličiny (tedy hodnota první fyzikální veličiny) a μ_y je skutečná hodnota 2. měřené náhodné veličiny (tedy hodnota druhé fyzikální veličiny). ϑ představuje předem zvolenou konstantu, jejímž vhodným zvolením můžeme upravovat testovanou hypotézu (nejčastěji volíme $\vartheta = 0$, což odpovídá shodným hodnotám měřených fyzikálních veličin).

Nyní potřebujeme odvodit podobu testové statistiky. V minulém dílu seriálu jsme se zabývali vícerozměrnou verzí centrální limitní věty, kterou nyní použijeme (pokud jste zapomněli, co vícerozměrná CLV je, bylo by dobré si to připomenout). Nebudeme zde úplně podrobně

[5] Rozdíl oproti t-testu je ten, že v tomto případě obě fyzikální veličiny měříme, zatímco v případě t-testu jsme měřili jen jednu fyzikální veličinu a porovnávali ji s předem známou konstantou.

[6] Ve skutečnosti potřebujeme, aby měřená data měla konečný rozptyl. Tento předpoklad se ovšem v praxi nikdy neověřuje a pokládá se za splněný, proto se jím v tomto textu nebudeme dále zabývat.

popisovat dosazení do vícerozměrné centrální věty, pouze uvedeme, že pokud zvolíme funkci f jako

$$f(x,y) = x - y\,,$$

dostaneme[7], že za platnosti hypotézy platí

$$Z_{n,m} = \frac{\overline{x_n} - \overline{y_m} - \vartheta}{\sqrt{\frac{S^2_{n_x}}{n} + \frac{S^2_{m_y}}{m}}} \xrightarrow{D} N(0,1)\,,$$

kde $S^2_{n_X}$ označuje výběrový rozptyl měření odpovídající první fyzikální veličině, $S^2_{m_Y}$ označuje výběrový rozptyl měření odpovídající druhé fyzikální veličině a $\overline{x_n}$, $\overline{y_m}$ označují příslušné výběrové průměry.

Naprosto stejnou úvahou jako v případě t-testu dojdeme k závěru, že za platnosti alternativy bude testová statistika s největší pravděpodobností nabývat velmi velkých kladných nebo záporných hodnot. Kritický obor z-testu je tedy optimální zvolit jako

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

kde u označuje kvantil normálního rozdělení $N(0,1)$. Takováto volba kritického oboru zajistí, že za platnosti hypotézy bude pravděpodobnost, že realizovaná hodnota testové statistiky bude ležet v kritickém oboru, právě α (tj. pravděpodobnost chyby 1. druhu bude za platnosti hypotézy rovna α). Zároveň bude pro velký počet měření (myšleno pro velké m i n) pravděpodobnost, že za platnosti alternativy padne realizovaná hodnota testové statistiky do kritického oboru, velká. Kritický obor je tedy zvolen optimálně.

Velice jednoduše by šel z-test upravit pro jednostrannou hypotézu a alternativu, tedy pro testování hypotézy a alternativy ve tvaru

$$H: \mu_x - \mu_y \geq \vartheta\,,$$
$$A: \mu_x - \mu_y < \vartheta\,.$$

Stačilo by použít stejný princip jako v případě t-testu. Pokud by byla znaménka nerovností opačná, také by nebyl problém z-test příslušně modifikovat.

χ^2 *test rozptylu*[8]

χ^2 test rozptylu použijeme ve chvíli, kdy chceme otestovat, že rozptyl měřených dat je roven nějaké předem známé konstantě σ_0^2. Hypotéza a alternativa budou tedy vypadat následovně

$$H: \sigma_x^2 = \sigma_0^2\,,$$
$$A: \sigma_x^2 \neq \sigma_0^2\,,$$

kde σ_x^2 představuje skutečnou hodnotu rozptylu našich měřených dat.

Pro tento test je důležitý předpoklad toho, že naše měřená data mají normální rozdělení (s libovolnými parametry μ a σ^2). Bez tohoto předpokladu tento test nefunguje[9] (ani asymptoticky) a nemůžeme ho tedy používat.

[7] Pozorný čtenář by si měl toto tvrzení detailně rozmyslet (sám odvodit). Je potřeba použít předpoklad o nezávislosti měřených dat.

[8] Toto řecké písmeno se vyslovuje [chi:], a název testu se obvykle vyslovuje jako χ kvadrát test.

[9] Ve většině případů v praxi je předpoklad o normalitě měřených dat oprávněný.

Jako testovou statistiku zvolíme následující transformaci měřených dat

$$CH_n = \frac{(n-1)S_n^2}{\sigma_0^2}\,,$$

kde S_n^2 je výběrový rozptyl a n je počet měření. Vzhledem k rozsahu tohoto textu není v našich silách podrobně vysvětlit, proč je zrovna tato volba testové statistiky nejvhodnější, ani nezvládneme podrobně odvodit, jaké má za platnosti hypotézy rozdělení. Na tomto místě proto jen uvedeme fakt, že testová statistika CH_n má za platnosti hypotézy rozdělení χ^2_{n-1} (čti χ kvadrát o $n-1$ stupních volnosti).

Rozdělení χ^2_k je dalším často se vyskytujícím absolutně spojitým rozdělením. Jeho hustota pravděpodobnosti má příliš komplikovaný tvar na to, abychom ji zde uváděli, také by to nemělo příliš smysl. Pro naše účely se spokojíme s tím, že si v tabulkách můžeme najít kvantily tohoto rozdělení, které budeme značit $\chi^2_k(\alpha)$, kde α značí hladinu ($\in (0,1)$) a k značí počet stupňů volnosti rozdělení. Doporučujeme každému čtenáři, aby si v matematickém softwaru vykreslil graf hustoty pravděpodobnosti rozdělení χ^2_k pro různé hodnoty stupňů volnosti k, aby získal představu o tom, jak takováto hustota vypadá a jak se mění v závislosti na počtu stupňů volnosti.

Nyní už k volbě kritického oboru. Lze snadno nahlédnout, že za platnosti hypotézy bude realizovaná hodnota testové statistiky s největší pravděpodobností blízká hodnotě $n-1$. To plyne z faktu, že výběrový rozptyl S_n^2 bude s největší pravděpodobností pro velký počet měření nabývat hodnot blízkých skutečnému rozptylu σ_0^2. Zlomek $\frac{S_n^2}{\sigma_0^2}$ tedy bude s největší pravděpodobností pro velký počet měření nabývat hodnot blízkých 1. Pokud bude platit alternativa, bude testová statistika nabývat buď výrazně menších hodnot než $n-1$ (v případě, že $\sigma^2 < \sigma_0^2$), nebo výrazně větších hodnot než $n-1$ (v případě, že $\sigma^2 > \sigma_0^2$). Kritický obor bude proto zvolen následovně[10]

$$C = \left(0, \chi^2_{n-1}\left(\frac{\alpha}{2}\right)\right) \cup \left(\chi^2_{n-1}\left(1-\frac{\alpha}{2}\right), \infty\right)\,.$$

Takto zvolený kritický obor bude mít opět všechny požadované vlastnosti.

Tento test by se opět dal velice jednoduše upravit pro případ jednostranné hypotézy a alternativy.

Několik poznámek ke konstrukci a interpretaci výsledků testů

Na konec uveďme několik krátkých poznámek.

- Pokud nějaký test zamítá hypotézu, můžeme tvrdit, že hypotéza nejspíš neplatí (z konstrukce testu víme, že pokud proneseme takovýto výrok, budeme se mýlit jen v α procentech případů). Musíme ale mít stále na paměti možnost, že jsme mohli udělat chybu 1. druhu, tedy nemůžeme nic najisto tvrdit!

- Pokud test nezamítá hypotézu, je interpretace výsledku ještě složitější. Buď to znamená, že hypotéza skutečně platí, nebo to znamená, že sice platí alternativa, ale naměřili jsme data, pro která hypotézu nezamítáme (udělali jsme tedy chybu 2. druhu). S chybou druhého druhu je to poněkud složitější, neboť nemáme pod kontrolou, jaká je pravděpodobnost

[10] Testová statistika může z definice nabývat jen kladných hodnot, proto nemá smysl do kritického oboru zařazovat záporné hodnoty.

chyby 2. druhu za platnosti alternativy. Správnou volbou kritického oboru můžeme zajistit, že za platnosti alternativy bude pravděpodobnost chyby 2. druhu co možná nejmenší a pro velký počet měření bude konvergovat k nule, ale jaká tato pravděpodobnost v konkrétním případě bude, nevíme[11].

Představme si například, že chceme t-testem otestovat hypotézu

$$\begin{aligned} H :&\mu_x = 0\,, \\ A :&\mu_x \neq 0\,, \end{aligned}$$

ale skutečná hodnota střední hodnoty měřených dat je například $\mu_x = 0{,}000\,000\,01$. Technicky bychom mohli říci, že platí alternativa, ale nemůžeme očekávat, že pokud nebudeme mít opravdu hodně měření, že nám t-test vyjde ve prospěch alternativy. Proto je nutné být velice opatrný při interpretaci výsledků statistických testů, zejména v případě, že výsledkem je nezamítnutí hypotézy.

- Postavení hypotézy a alternativy není rovnocenné. Pokud je to možné, je vždy lepší volit za alternativu to tvrzení, které bychom chtěli potvrdit. Jak je psáno výše, kde je rozebírána interpretace výsledků testů, v případě zamítnutí hypotézy potom máme větší právo tvrdit, že jsme opravdu potvrdili správnost našeho tvrzení (neboť jsme přesvědčivě vyvrátili opačné tvrzení). Je nutné poznamenat, že vzhledem ke konstrukci testů nelze takovéto možnosti vždy dosáhnout, potom se musíme spokojit s opačnou volbou hypotézy a alternativy.

- Volba hladiny testu α je velice důležitá a je nutné ji provést před provedením každého testu. Její volba závisí na tom, jak chceme mít vyváženou pravděpodobnost chyby 1. a 2. druhu. Hladina spolehlivosti testu α je přímo rovna pravděpodobnosti chyby 1. druhu. Pokud hladinu testu zvolíme malou, budeme mít malou pravděpodobnost chyby 1. druhu, ale naopak větší pravděpodobnost chyby 2. druhu. To souvisí s tím, že pro malé hodnoty α bude kritický obor každého testu menší než pro velké hodnoty α (rozmyslete si sami). Obvykle se volí hladina spolehlivosti rovna $\alpha = 0{,}05$. Pokud děláme nějaký důležitý experiment, kde opravdu záleží na správnosti závěrů, snažíme se pravděpodobnost chyby 1. druhu co nejvíce omezit, volíme tedy menší hladinu spolehlivosti (typicky $\alpha = 0{,}01$ nebo $\alpha = 0{,}005$). Při řešení seriálové úlohy pracujte s hladinou spolehlivosti $\alpha = 0{,}05$.

Zamítání hypotézy na základě p-hodnoty

Doteď jsme rozhodovali o zamítání hypotézy na základě realizované hodnoty testové statistiky a kritického oboru (pokud realizovaná hodnota testové statistiky ležela v kritickém oboru, zamítali jsme hypotézu). Tento postup je velice intuitivní a snadno představitelný, ale jeho nevýhoda je, že máme jen malou informaci o tom, zda jsme hypotézu (ne)zamítali s rezervou, nebo zda se pohybujeme na hranici mezi zamítáním a nezamítáním hypotézy. Proto se pro každý statistický test a konkrétní naměřená data definuje tzv. p-hodnota, která slouží jako alternaiva k rozhodování o (ne)zamítání hypotézy. p-hodnotu testu pro konkrétní naměřená data definujeme jako takové p, že realizovaná hodnota testové statistiky leží právě na hranici kritického oboru pro hladinu p. p-hodnotu lze také interpretovat jako pravděpodobnost, že

[11] Ve skutečnosti ve většině případů je možné pravděpodobnost chyby 2. druhu vyčíslit, ale je to příliš náročné na to, abychom se tím podrobněji zaobírali.

bychom za platnosti hypotézy naměřili data, která protiřečí hypotéze ještě více než ta data, která máme naměřená. Rozmyslete si sami, že definice p-hodnoty a její interpretace dávají smysl.

Rozhodování o zamítnutí hypotézy na základě p-hodnoty je jednoduché, stačí porovnat p-hodnotu se zvolenou hladinou testu α. Pokud je p-hodnota menší než hladina testu α, zamítáme hypotézu, neboť to znamená, že jsme naměřili (za platnosti hypotézy) opravdu extrémní data.

Většina moderních matematických programů (včetně R, které používáme) preferuje rozhodování na základě p-hodnoty, kterou poskytuje jako výstup. Naopak kritický obor většinou nebývá ve standardním výstupu matematických programů uváděn.

Další statistické testy

V tomto díle seriálu jsme uvedli jen několik základních typů statistických testů, ve skutečnosti jich existuje mnohem více. Pokud během fyzikálních experimentů narazíte na nějakou hypotézu, kterou potřebujete pomocí naměřených dat otestovat, je velice pravděpodobné, že na to bude existovat speciální statistický test. Stačí si ho vyhledat na internetu. Pro použití statistického testu v praxi ani není nutné znát přesnou podobu testové statistiky, kritického oboru ani přesně vědět, jak tento test funguje, neboť tyto věci za nás většinou spočte matematický software. Jediné, čemu je opravdu potřeba rozumět, je hlavní idea testování hypotéz, správně pochopit předpoklady tohoto testu a umět interpretovat výsledky (zejména na základě p-hodnoty). Pro informaci zde ještě uvedeme několik statistických testů včetně jejich předpokladů. Pokud byste některý chtěli v praxi použít, stačí si ohlídat splnění všech jeho předpokladů a umět interpretovat jeho výsledky na základě p-hodnoty.

χ^2 *test dobré shody rozdělení*

Tento test se použije v případě, že chceme otestovat, zda měřená data pocházejí z určitého diskrétního rozdělení se známými parametry (je důležité, aby toto rozdělení bylo diskrétní, v případě spojitého rozdělení by se použil Kolmogorovův-Smirnovův test, který popíšeme později). Předpokládejme, že máme nějaké známé diskrétní rozdělení R. Pomocí χ^2 testu dobré shody rozdělení můžeme otestovat následující hypotézu a alternativu

$$H : \text{měřená data mají rozdělení } R\,,$$
$$A : \text{měřená data nemají rozdělení } R\,.$$

Pokud bychom neznali všechny parametry rozdělení R (například bychom věděli, že se jedná o Poissonovo rozdělení, ale neznali bychom parametr λ), můžeme si tyto neznámé parametry odhadnout z naměřených dat a použít je v našem testu, který bude fungovat i po této drobné modifikaci.

Kolmogorovův-Smirnovův test

Tento test se použije v případě, že chceme otestovat, zda měřená data pocházejí z určitého spojitého rozdělení se známými parametry (je důležité, aby toto rozdělení bylo spojité, v opačném případě musíme použít χ^2 test dobré shody rozdělení). Předpokládejme, že máme nějaké známé spojité rozdělení S. Pomocí Kolmogorovova-Smirnovova testu můžeme otestovat následující hypotézu a alternativu

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

Opět platí, že pokud neznáme všechny parametry rozdělení S, můžeme je odhadnout z naměřených dat a tento test bude stále fungovat.

Dvouvýběrový Kolmogorovův-Smirnovův test

Tento test předpokládá, že máme 2 nezávislé sady měření, které obě pocházejí z nějakých spojitých rozdělení, a chceme otestovat hypotézu, že obě tyto sady měření mají stejné rozdělení. Pokud označíme teoretické rozdělení první sady měřených dat jako X a teoretické rozdělení druhé sady dat jako Y, potom dvouvýběrový Kolmogorovův-Smirnovův test testuje následující hypotézu a alternativu

$$\begin{aligned} H &: X \text{ a } Y \text{ mají stejná rozdělení}\,,\\ A &: X \text{ a } Y \text{ nemají stejná rozdělení}\,. \end{aligned}$$

Test korelačního koeficientu

Tento test použijeme ve chvíli, kdy máme 2 nezávislé sady měření, které mají obě normální rozdělení, a chceme otestovat, zda korelace mezi nimi je nulová nebo nikoliv. Pokud označíme jako X teoretické rozdělení první sady měření, jako Y teoretické rozdělení druhé sady měření a jako ϱ označíme jejich korelační koeficient, tedy

$$\varrho = \mathrm{corr}(X, Y)\,,$$

potom test korelačního koeficientu testuje následující hypotézu oproti alternativě

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

Jen poznamenejme, že pro tento test je opravdu podstatný předpoklad o normálním rozdělení a nezávislosti měřených dat.

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.