Úvod a možná motivace

V současnosti atraktivní oblastí užití „prostředků“ testování statistických hypotéz je biomedicína. V ní se často užívá pojmů „Senzitivita a specificita biomedicínských testů“. K jejich zavedení viz např. https://portal.matematickabiologie.cz/index.php?pg=aplikovana-analyza-klinickych-a-biologickych-dat--biostatistika-pro-matematickou-biologii--vztah-pravdepodobnosti-statistiky-a-biostatistiky--senzitivita-specificita-a-prediktivni-hodnoty .[1]

Z této poznámky a obsahu odkazu je zřejmé: statistická zkoumání **MUSÍ** dělat osoba (častěji, komunikující a komunikativní skupina osob = lidí obou odborností, kteří se chtějí a umějí domluvit) výborně zvládající oblast, odkud jsou získána data a oblast statistických metodik, a to obé do značné hloubky. Jinak je produkt velmi pochybné úrovně (nejčastěji „ovzorečkovaný“ blábol s nějakými výpočty, v současnosti je to velmi frekventované).

Námět: „Přeložte“ pojmy „Senzitivita a specificita testů“ do pojmů užívaných ve statistickém testování.

Vlastní *poznámky* k praktickému používání „klasických statistických testů hypotéz“

1. Hypotézu a alternativu nelze volit libovolně, vyplývají z řešeného problému. Tj. jak pro hypotézu, tak i pro alternativu **musí být důvody z (odborné, vědecké, …) oblasti odkud přicházejí data**. [2]

2. Řešené úlohy jsou dichotomií, „vybíráme“ hypotézu nebo (vylučovací≡XOR) alternativu. Nelze vybrat nic jiného. Klasický statistický test musí vždy rozhodnout (na rozdíl např. od sekvenčních testů, …). Rozhodnutí testu musí být vždy jen a jen pro hypotézu nebo (vylučovací XOR) pro alternativu. Nelze připustit současnou platnost hypotézy a alternativy. V „klasickém statistickém testu“ není přijatelné, aby (pro zkoumaná data) platilo něco jiného, než buď hypotéza nebo (vylučovací XOR) alternativa.

3. Z povahy postupů pro řešení „klasického statistického testu“ plyne, že platnost hypotézy je preferována. Alternativa slouží jako možnost, která by mohla platit, pokud dojde (testem, na základě disponibilních dat) k zamítnutí hypotézy. Alternativa je tedy jakousi „nouzovou“ variantou k hypotéze. **Toto musí být zřejmé z řešeného problému.** Tedy důvody jsou mimo statistiku.
Je dobré si uvědomit, že dojde-li k záměně alternativy za hypotézu a hypotézy za alternativu nebude příslušný test „pouhým otočením“ původního testu!

---

[1] K tam uváděným symbolům a, b, c, d viz „**Tab. 2.1:** Sumarizace výsledků ultrazvukového vyšetření jater vzhledem k laboratornímu ověření.“. Z ní a hlavně z příkladu pod ní, lze jednoznačně pochopit význam těchto symbolů. Je dobré si uvědomit, že tam uváděné hodnoty jsou hodnotami VÝBĚROVÝMI. Striktně řečeno jedná se o bodové odhady (= spočítané z pozorovaných dat) uvedených pojmů (abstraktních). To však neoslabuje pochopitelnost zaváděných pojmů.

[2] Zde je dobré si uvědomit, že ne pro každou dvojici hypotéza – alternativa existuje nebo je dostupný nebo je použitelný test. „Statistika je mlčky skromná, protože toho zas tak moc neumí“. Proto je často třeba k přeformulování nebo i doplnění nebo zúžení původního problému.

Velice volná, nepřesná ale názorná interpretace: Věříme (podloženě, argumenty z oblasti odkud pocházejí data), že hypotéza platí a od „napozorovaných“ dat očekáváme, že tuto víru nenaruší. Pokud data tu naší důvěru nepotvrdí (ve statistickém testu), kloníme se k alternativě.

4. Klasický statistický test předpokládá data ve formě **náhodného výběru** (iid) pevného **nenáhodného** rozsahu.

Formulace úlohy „klasického testu hypotézy proti alternativě“

Mějme náhodný výběr $\{x_1, x_2, ..., x_n\}$ testujeme hypotézu H proti alternativě A, k testu hypotézy použijeme statistiku $T(x_1, x_2, ..., x_n)$, která bude nazývána testovým kritériem. **Celý** obor hodnot, kterých může statistika $T(x_1, x_2, ..., x_n)$ nabývat, rozdělíme na dvě **disjunktní části** W a její doplněk. Pokud hodnota testového kritéria padne do W zamítáme hypotézu H ve prospěch alternativy A. Obor W bude nazýván kritickým oborem. Pokud hodnota T, testového kritéria padne mimo kritický obor W, testovanou hypotézu nezamítáme.

Ve statistice (přesněji v klasickém přístupu k statistickému testování) je hypotéza formulována tak, že rozdělení pravděpodobnosti zkoumaných dat patří nějaké množině pravděpodobnostních rozdělení $\mathcal{P}_H$ a alternativa je formulována tak, že rozdělení pravděpodobnosti zkoumaných dat patří nějaké množině pravděpodobnostních rozdělení $\mathcal{P}_A$. Nic jiného nastat nemůže a navíc $\mathcal{P}_H \cap \mathcal{P}_A = \emptyset$, $\mathcal{P}_H \neq \emptyset$ a $\mathcal{P}_A \neq \emptyset$[3]. Je pak na statistikovi (ve spolupráci s odborníkem na oblast odkud pocházejí data, pokud jím sám statistik není), aby formulace z oblasti odkud data pocházejí do těchto pojmů „přeložil“.

Pokud je $\mathcal{P}_H$ resp. $\mathcal{P}_A$ jednoprvková mluvíme o jednoduché hypotéze, resp. alternativě. Pokud je $\mathcal{P}_H$ nebo $\mathcal{P}_A$ více prvková mluvíme o složené hypotéze nebo alternativě.

Někdy se místo názvu hypotéza používá název nulová hypotéza, značeno H_0, někdy se také značí alternativa H_1.

Chyby takového rozhodování

- **Chyba prvního druhu:** Hypotéza H **platí,** ale **je** na základě testového kritéria **zamítnuta** $T(x_1, x_2, ..., x_n) \in W$.
- **Chyba druhého druhu:** Hypotéza H **neplatí,** ale **není** na základě testového kritéria **zamítnuta** $T(x_1, x_2, ..., x_n) \notin W$.

Pravděpodobnosti chyb

Pravděpodobnost chyby prvního druhu: $\alpha = P(T \in W / H)$. α je pravděpodobností chyby prvního druhu = hladinou významnosti testu = velikostí kritického oboru = … . Běžně bývá zadána tato pravděpodobnost a k ní se určuje odpovídající kritický obor. Taková skutečnost bývá značena $\alpha = P(T \in W_\alpha / H)$. Kritický obor je pak řešením této

[3] Zcela výjimečně může být jedna z obou množin prázdná. Jednalo by se o testy, které by nezávisle na „napozorovaných datech“ trvale nezamítaly hypotézu nebo (vylučovací = XOR) trvale zamítaly hypotézu ve prospěch alternativy.

rovnice k danému α vůči W_α. Pro složené hypotézy lze zavést kritický obor následující nerovností $\alpha \geq P_{P\in H}(T \in W_\alpha / H)$. Danému α je pak přiřazen „maximální" kritický obor vyhovující $\alpha \geq P_{P\in H}(T \in W_\alpha / H); \forall P \in H$.

p-hodnota testu: je nejmenší hodnotou hladiny významnosti, při konkrétním testu, při které by bylo možno ještě H zamítnout. Detaily viz např.: https://www.apps.stat.vt.edu/leman/VTCourses/schervish-pvals.pdf . V některých modifikacích (vázaných na konstrukci testu) může být p-hodnota formulována jako hodnota hladiny významnosti na které se mění rozhodnutí ze zamítnutí hypotézy ve prospěch alternativy na rozhodnutí o nezamítnutí hypotézy.

Pravděpodobnost chyby druhého druhu: $1-\beta = P(T \notin W_\alpha / A) = 1 - P(T \in W_\alpha / A)$. Pravděpodobnost β je pak nazývána silou testu založeného na kritickém oboru W_α nebo přímo silou kritického oboru W_α. Síla kritického oboru (testu) je pravděpodobností zamítnutí hypotézy za platnosti alternativy.

Protože platí to, že s růstem pravděpodobnosti chyby prvního druhu obvykle klesá (přesněji neroste) pravděpodobnost chyby druhého druhu a naopak (viz přesnou formulaci a důkaz v příloze 1. tohoto textu).

Odtud plyne **klasický přístup k testování**, zvolí se horní mez chyby 1-ho druhu $0 < \alpha < 1$ a hledá se takový test (takové testové kritérium, kritický obor W_α), který minimalizuje chybu druhého druhu $1-\beta_{W_\alpha}$ (= maximalizuje sílu testu β_{W_α}). Toto je zřejmé, pokud je alternativa jednoduchá. Pokud je složená, pak hledáme takový test, který, při daném omezení na pravděpodobnost chyby prvního druhu, maximalizuje sílu testu pro „každý prvek" alternativy (samozřejmě pokud takový test existuje).

Pro základy možných (a i variantních) postupů řešení klasické úlohy „testování hypotéz" viz např.: Jaroslav Hátle, Jiří Likeš: Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974, str. 277-361 nebo jakákoliv trochu solidnější učebnice statistky obsahující partie „Testování hypotéz".

Užívání *p-hodnot* testů při daných datech

p-hodnota je svojí povahou efektivní pojem umožňující srovnávání výsledků různorodých testů na různorodých datech. Je však zapotřebí uvědomit si, že tvrzení „čím nižší dosažená p-hodnota, tím je výsledek testu lepší" nemusí být úplně korektní. Jak bylo výše uvedeno klesající pravděpodobnost chyby prvního druhu znamená (až na velmi drobné výjimky) rostoucí pravděpodobnost chyby druhého druhu. Reálně (mimo statistiku, v oblasti odkud pocházejí data) může být pravděpodobnost chyby druhého druhu – i když minimalizovaná konstrukcí testu – ještě dost velká na to, aby byla přijatelná (přípustná) pro aplikovatelnost takových statistických výsledků. Proto je, současně s p-hodnotou, důležitá dosahovaná hodnota síly testu nebo alespoň nějaké její vymezení. Bez toho může být (a velmi často bývá) „velice dobrá" p-hodnota pouhým „plácnutím do vody". Pozornost je též třeba věnovat postupům, pomocí nichž je p-hodnota počítána. Také pozor na předpoklady takového výpočtu.

Příloha 1.

Mějme náhodný výběr $\{x_1, x_2, ..., x_n\}$ a statistiku $T(x_1, x_2, ..., x_n)$. Uvažujeme kritický obor W_α, pro který je dle definice $\alpha \geq P(T \in W_\alpha / H)$, dále další kritický obor $W \subset W_\alpha$, takový že: $P(T \in W_\alpha - W / H) > 0; P(T \in W_\alpha - W / A) > 0$.

Potom:

$\alpha_W = P(T \in W / H) = P(T \in W_\alpha / H) - P(T \in W_\alpha - W / H) < P(T \in W_\alpha / H)$ a

$1 - \beta_W = 1 - P(T \in W / A) = 1 - (P(T \in W_\alpha / A) - P(T \in W_\alpha - W / A)) =$

$= 1 - P(T \in W_\alpha / A) + P(T \in W_\alpha - W / A) > 1 - \beta_{W_\alpha}$

Opačně: Mějme kritický obor $W \supset W_\alpha$, takový že: $P(T \in W - W_\alpha / H) > 0; P(T \in W - W_\alpha / A) > 0$.

Potom: $\alpha_W = P(T \in W / H) = P(T \in W - W_\alpha / H) + P(T \in W_\alpha / H) > P(T \in W_\alpha / H)$ a

$1 - \beta_W = 1 - P(T \in W / A) = 1 - (P(T \in W_\alpha / A) + P(T \in W - W_\alpha / A)) =$

$= 1 - P(T \in W_\alpha / A) - P(T \in W - W_\alpha / A) < 1 - \beta_{W_\alpha}$