



StatSoft

Síla testu

Cílem tohoto článku je uvedení do problematiky výpočtu síly testu s využitím modulu *Analýza síly testu* ve *STATISTICA*.

Testování hypotéz

Každý statistický test je ve své podstatě rozhodovací pravidlo, které pomáhá určit, zda je zkoumaná hypotéza pravdivá, či nikoli. Na základě nasbíraných dat (za splnění předpokladů konkrétního testu) je vyhodnocena **testová statistika** a podle její hodnoty se určí, zda platí či neplatí tzv. **nulová hypotéza** pro celou populaci reprezentované vzorkem dat. Vybíráme zpravidla takový test, který uvádí jako **nulovou hypotézu** opačné tvrzení, než jaké máme v úmyslu dokázat. To proto, že testy se snaží vždy nasbírat důkazy o tom, že nulová hypotéza neplatí. Když se to podaří, nulovou hypotézu zamítáme, a je tak prokázáno tvrzení, které jsme dokázat chtěli. Závěry testů jsou platné vždy jen s určitou



pravděpodobností, obvykle volíme **hladinu spolehlivosti testu 5 %**. To znamená, že pokud bychom měli šanci provést 100 náhodných výběrů z téže populace, o níž víme, že ve skutečnosti platí nulová hypotéza, aplikací testu na každý z nich získáme maximálně 5 případů, kdy nulovou hypotézu zamítneme. V případě, že zamítáme nulovou hypotézu, je tedy možné, že tak činíme nesprávné rozhodnutí způsobené použitím nešťastně zvoleného vzorku dat – pravděpodobnost, že tato **chyba prvního druhu** nastane, je ovšem pouze 5 %. Více o principu testování hypotéz se lze dočíst v našem článku [Pohádkový soudní proces](#).

Síla testu

Požadavkem na hladinu spolehlivosti testu omezujeme pravděpodobnost pro chybu prvního druhu (*nulová hypotéza platí a je přesto zamítnuta*), nicméně může nastat též situace, kdy dojde k **chybě druhého druhu**; té se dopouštíme tehdy, když nulová hypotéza ve skutečnosti neplatí, ale k jejímu zamítnutí nedojde. **Síla testu** je číslo mezi 0 a 1, které udává pravděpodobnost, že při neplatnosti nulové hypotézy dojde k jejímu zamítnutí, tedy pravděpodobnost odhalení neplatnosti nulové hypotézy. Platí, že čím vyšší je síla testu, tím lépe.

Je možné, že jste se již setkali s tvrzeními, které říkají: tento test je silnější než jiný pro nějakou alternativu. To úzce souvisí s právě definovaným pojmem. Test je silnější pokud při stejné pravděpodobnosti chyby prvního druhu, má větší sílu testu.

Proč je síla testu důležitá?

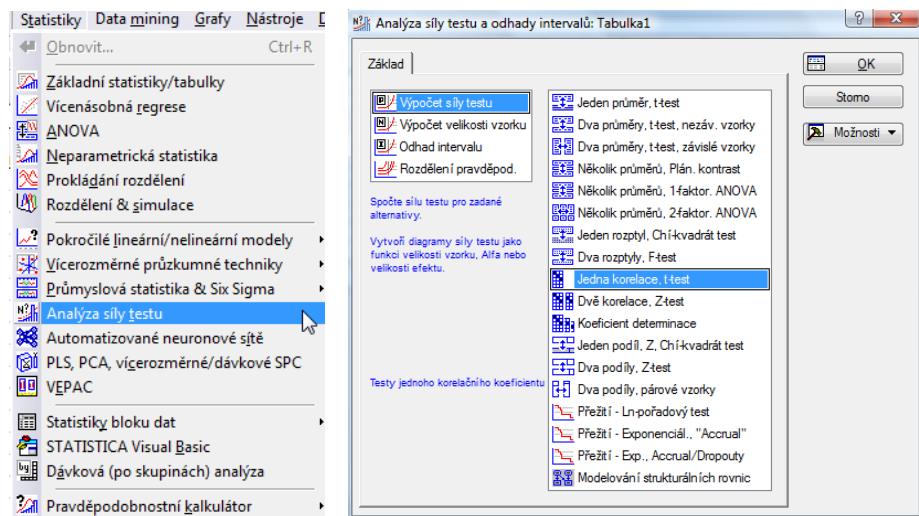
Síla testu je důležitá zejména tehdy, kdy nedojde k zamítnutí nulové hypotézy. Pokud je síla testu malá, je oprávněně zamítnutí nulové hypotézy málo pravděpodobné a je tedy velice malá šance prokázat zkoumanou hypotézu, která je protikladem nulové hypotézy. Doporučovaná minimální hodnota pro sílu testu je 0.7 (chyba druhého druhu má tak pravděpodobnost maximálně 30 %). Sílu testu ovlivňuje několik parametrů: typ zvoleného testu, vzorek dat reprezentující populaci a zejména také **velikost vzorku dat**. U vědeckých studií je proto dobrým zvykem, s přihlédnutím k předpokládaným metodám zpracování dat, předem zvolit rozsah výběru tak, aby byla zaručena požadovaná minimální síla testu.

Příklad: Test nulovosti korelačního koeficientu

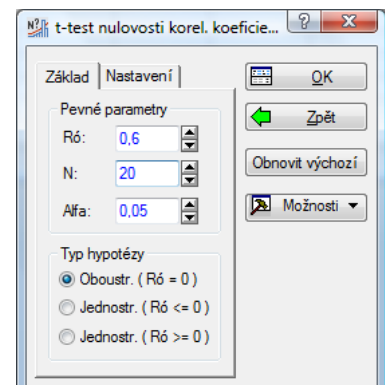
Často řešenou praktickou otázkou je závislost dvou spojitých veličin. Lineární závislost posuzujeme na základě **Pearsonova korelačního koeficientu**, je-li jeho hodnota blízká nule, není mezi veličinami lineární závislost, blíží-li se 1 (maximální hodnota), naznačuje přímou úměrnost - když roste jedna veličina, roste spolu s ní lineárně i druhá, blíží-li se hodnotě -1 (minimální hodnota), pak jedna veličina klesá s tím, když druhá roste. Při výpočtu korelace stanovujeme i její statistickou významnost na základě testu nulovosti korelačního koeficientu. Korelace je signifikantní, jestliže je testem nulovost korelace zamítnuta.

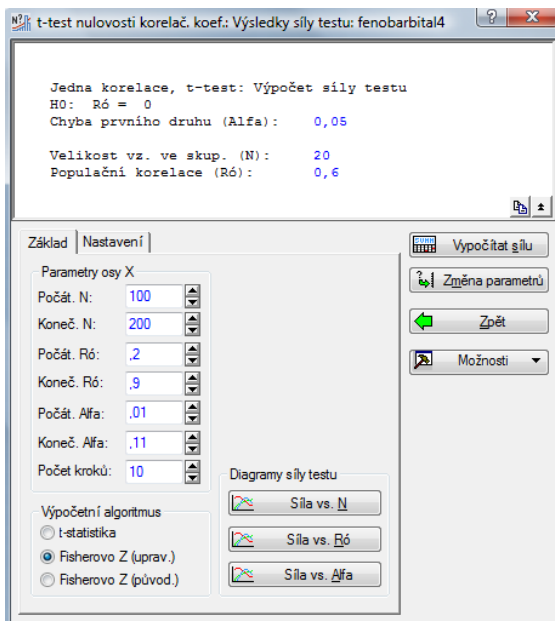
Ukážeme na jednoduchém příkladu, jaký vliv má na sílu testu rozsah vzorku dat a hodnota výběrového korelačního koeficientu. Předpokládejme, že hodnota výběrové korelace dvou veličin je 0.6, a hodnota korelace byla vypočtena na základě 20 měření dvou veličin, podle oboustranné p -hodnoty testu nulovosti korelace $p=0.0042$ je tato hodnota statisticky významná. Jaká je síla testu nulovosti korelace pro rozsah výběru $N=20$, a jak se bude měnit, jestliže bude rozsah menší anebo větší?

Pro výpočet síly testu zvolíme menu **Statistiky – Analýza síly testu**. V levém menu následujícího okna vybereme **Výpočet síly testu** a v pravém příslušný test: **Jedna korelace, t-test**. Potvrdíme kliknutím na **OK**.

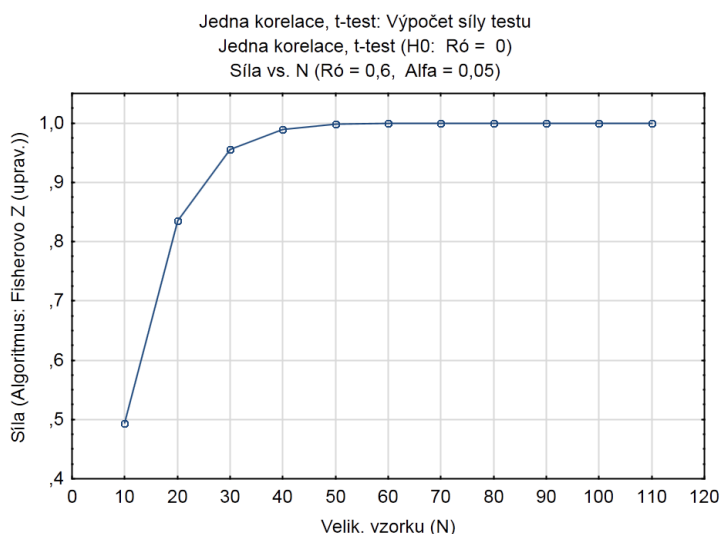


V dalším kroku zadáme vypočtenou hodnotu výběrové korelace jako parametr R_0 , rozsah výběru N a dále použitou hladinu α pro chybu prvního druhu. Typ hypotézy volíme podle typu alternativní hypotézy (oboustranná či jednostranná), opět potvrdíme **OK**.

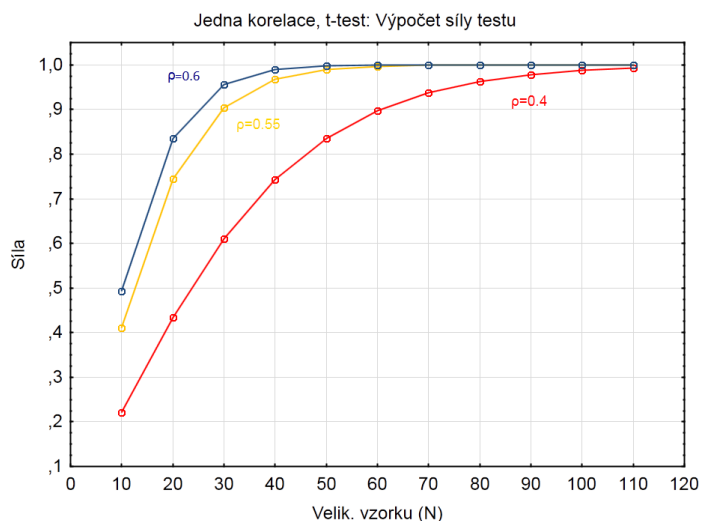




Tlačítkem *Vypočítat sílu* získáme výslednou sílu testu, která je 0.8351. Shrnutí pro různé rozsahy výběru získáme nejlépe z diagramu síly testu *Síla vs. N*, přičemž parametry grafu *Počáteční N* a *Konečné N* nastavíme od 10 do 110. Křivka síly testu ukazuje, že pro deset pozorování je síla testu nízká (asi 0.5), pro dvacet pozorování je síla testu již vyšší než 0.8 a velmi blízká jedné je pro vzorky o rozsahu alespoň 40 případů. Zvýšením rozsahu výběru tedy lze zvyšovat sílu testu, ale v rozumné míře – náklady na získání nadměrně velkého počtu dat totiž přinesou jen nepatrné zlepšení (pro korelaci 0.6 již 20 pozorování zajišťuje sílu větší než 0.8, volit v tomto případě rozsah datového souboru např. 70 je proto zbytečné).



Pokud by výběrová korelace byla nižší, například 0.4, není již při rozsahu výběru $N=20$ statisticky významná, a i síla testu se změní – dosahuje pouze hodnoty 0.4342. Graf závislosti síly testu na velikosti vzorku získáme analogicky jako pro korelaci 0.6. Sloučením grafů a nastavením odlišných barev křivek síly testu získáváme přehledný graf demonstrující závislost síly testu na hodnotě výběrového korelačního koeficientu (ukázka konkrétně pro trojici hodnot 0.4, 0.55 a 0.6) a současně i velikosti vzorku (pro N od 10 do 110). Čím je absolutní hodnota korelace nižší, tím větší množství dat potřebujeme k dosažení stejné síly testu jako pro vyšší absolutní hodnoty korelace. Každá z křivek síly testu roste nejprve strmě a od určitého n roste už jen mírně. Rozsah výběru by měl být dostatečný, ale ne zbytečně velký. Jestliže předpokládáme, že korelace obou veličin nabývá ve skutečnosti hodnoty mezi 0.4 a 0.6, je vhodné zvolit rozsah výběru alespoň 40 (tj. takový, kdy i nejnižše položená křivka síly testu dosahuje síly alespoň 0.7).



Závěr

Modul *Analýza síly testu* ve *STATISTICA* poskytuje nástroje pro určení síly již provedených testů či stanovení optimálního rozsahu výběru pro zajištění dostatečné síly testů zamýšlených. Toto je velmi důležité a jedno z hlavních použití. Určení velikosti vzorku před samotnou studií může předejít mnoha problémům, především tedy s nedostatečnou silou pro prokázání cíle studie. Je jistě velká škoda provést nákladný výzkum a zjistit, že výsledek je nevyhovující, protože jsme si zvolili malý rozsah pro studii. Aby to nebylo tak jednoduché, může také nastat i druhá situace, představte si, že děláte studii, například test nějaké látky na zvířatech, předtím než látku aplikujete, nicméně musíte zvířata něčím nakazit. Do studie jste vzali mnoho pokusných zvířat, máte radost, že jste prokázali pozitivní účinky látky. Nicméně pak mohou přijít otázky opačného typu jako například: opravdu jste museli vystavovat tolik zvířat tak nebezpečné nemoci, nestačilo by jich méně? Bez analýzy síly testu budete jen těžko hledat argumenty.

Nadto lze pomocí tohoto modulu získat i užitečné intervalové odhady parametrů, které jste ve *STATISTICA* možná hledali, a je zde přístup i ke kalkulátoru pro vybraná zobecněná (necentrální) pravděpodobnostní rozdělení. Chcete-li se dozvědět o tématu analýza síly testu více, neváhejte navštívit náš [kurz Analýza síly testu](#).