



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Stanovení velikosti výzkumného souboru

Petr Soukup



2021

Informace o autorech:

Petr Soukup

FSV UK

soukup@fsv.cuni.cz



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

**MS
MT**
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Obsah

1	JAK URČIT VELIKOST VÝZKUMNÉHO SOUBORU?	4
2	NA ČEM ZÁLEŽÍ POŽADAVEK NA VELIKOST SOUBORU?.....	6
3	STANOVENÍ VELIKOSTI VÝZKUMNÉHO SOUBORU DLE POŽADOVANÉ CHYBY (PŘÍSTUP ZALOŽENÝ NA ŠÍŘI INTERVALŮ SPOLEHLIVOSTI).....	8
3.1	POSTUP PRO ODHAD STŘEDNÍ HODNOTY	8
3.2	POSTUP PRO ODHAD PROPORCE	10
4	PŘÍSTUP ZALOŽENÝ NA MINIMÁLNÍ POŽADOVANÉ SÍLE TESTU	13
4.1	STANOVENÍ VELIKOSTI SOUBORU PRO KORELACE	15
4.2	STANOVENÍ VELIKOSTI SOUBORU PRO ROZDÍL STŘEDNÍCH HODNOT (T-TESTY)	17
I	SHRNUTÍ	23
II	SEZNAM POUŽITÉ LITERATURY	24





1 JAK URČIT VELIKOST VÝZKUMNÉHO SOUBORU?

Tuto otázku si klade před provedením výzkumu snad každý výzkumník. Odpovědi, které se nabízí jsou v zásadě dvou druhů:

- a) Založené na zvyku či intuici
- b) Založené na statistických koncepcích.

Ad a) Na mnoha pracovištích a v jednotlivých oborech se tradují různá doporučení, která se přenášejí z generace na generaci. Jmenujme například doporučení typu „Udělej dvě stovky, tak do děláme vždy“, nebo „Když se počítá faktorová analýza, potřebuješ minimálně dvě stě jednotek“. Jak praví česká lidová pranostika „Na každém šprochu pravdy trochu“ nejsou tato doporučení úplně hloupá, ale jejich zdánlivá univerzálnost není reálná. Pro každou situaci (výzkumný problém, statistickou techniku, dělení do různých skupin) je potřebný jiný počet jednotek a je vhodné jej stanovit exaktně, aby následné výsledky analýz nebyly artefakty dat, ale aby šlo o seriózní a vědecky podložený výsledek. Proto přejdeme k postupům, které byly odvozeny v rámci statistiky.

Ad b) Tyto postupy nejsou jednotné a pracovní bychom je mohli rozdělit na dva základní přístupy:

- 1) Přístup založený na požadované velikosti chyby našich odhadů (šíři intervalů spolehlivosti)
- 2) Přístup založený na minimální požadované síle testu.

Bylo by možné ještě uvést další typy, ale dva uvedené jsou zcela jistě v praxi převažující.¹ Nutno pro úplnost dodat, že druhý uvedený postup je z výpočetního hlediska ještě rozdělen na dvě varianty, první je založená na přímých výpočtech skrze vzorce (lze tedy v případě nouze vypočítat ručně), druhá pak na simulacích Monte Carlo (zde je ruční výpočet nemyslitelný). Postupně projdeme jednotlivé varianty a na ukázkách ručních výpočtů i výpočtů ve statistickém software je budeme demonstrovat. Pro zjednodušení aplikace se omezíme

¹ Nutno dodat, že i když k tomu není žádná větší empirická studie, v praxi převažuje druhý postup, tj. užívání síly testu.





na software, který je zdarma a volně dostupný, o komerčním software bude referováno jen doplňkově.



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

**MS
MT**
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



2 NA ČEM ZÁLEŽÍ POŽADAVEK NA VELIKOST SOUBORU?

Než začneme rozebírat jednotlivé postupy stanovení velikosti výzkumného souboru, bude vhodné si obecně popsat, jaké charakteristiky ovlivňují vhodnou velikost. Jako první pravidla napadne analytika **variabilita dat**, tj. čím jsou data variabilnější, tím více jednotek bude pro zpracování potřeba a naopak u homogenních souborů je potřeba méně jednotek. Tato intuice je správná a bude potvrzena i v další části. Doplňme, že pro kategoriální data (a odhad proporce, resp. relativní četnosti) je variabilita proměnné funkcí příslušné proporce. Pro jevy málo a extrémně četné je příslušná variabilita nízká a pro jevy s proporcí okolo 50 % je naopak variabilita velká (srov. dále vzorce v další části). Tento zdánlivý paradox je nutné respektovat a pro jevy řídké, resp. extrémně časté, stačí menší soubory, než pro jevy, které se vyskytují s cca 50% proporcí.

Další faktory už tak zřejmé nebývají. U přístupu založeném na intervalech spolehlivosti (jejich šíři, srov. dále) je důležitým atributem právě **požadovaná šíře intervalu (přesnost odhadu)**. Za jinak stejných okolností záleží na požadované šířce intervalu spolehlivosti. Při požadavku na užší interval (přesnější odhad) potřebujeme tedy více jednotek, při požadavku na širší interval (méně přesný odhad) pak stačí méně výzkumných jednotek. Jak ukáží následné vzorce a na ně navazující výpočty, situace je pro výzkumníka nepříjemná v tom, že 2x větší požadavek na přesnost znamená 4x požadavek na velikost (vztah mezi přesností a velikostí tedy není lineární, ale je kvadratický).

Dalším faktorem, který ovlivňuje velikost souboru je **statistická technika**, kterou budeme chtít pro zpracování našich dat použít. Může to být překvapivé, ale už na počátku přemýšlení o výzkumném projektu bychom kromě výzkumných otázek a určení jednotek, které budeme zkoumat, měli mít poměrně jasno v tom, jaké statistické techniky budeme používat pro zpracování. Zjednodušeně platí pravidlo, že složitější technika vyžaduje ceteris paribus větší výzkumný soubor. Složitost techniky se odvozuje zejména od počtu proměnných, se kterými se zároveň pracuje.

Posledním faktorem, který ovlivňuje požadavek na velikost výzkumného souboru je úroveň efektu, který budeme zkoumat. Náznak tohoto faktoru jsme už uvedli výše u kategoriálních dat, nyní jej trochu rozvedeme a precizujeme. Představme si pro jednoduchost, že chceme zkoumat rozdíl klasické výuky matematiky a Hejného metody. Pro čistotu zvolíme





experimentální design a budeme děti náhodně rozdělovat do kontrolní skupiny (klasický přístup výuky) a experimentální (Hejného metoda). Intuitivně lze očekávat, že pokud je rozdíl (měřený například testem po roce výuky) mezi oběma metodami výuky velký, bude stačit poměrně malý výzkumný soubor, pokud je rozdíl nepatrný, bude naopak zapotřebí mít k dispozici velké množství zkoumaných žáků. U tohoto faktoru se ještě chvíli zastavíme. Zatímco určit na počátku úvah o výzkumném projektu, jaké budou požadované šíře intervalů spolehlivosti či jakou statistickou techniku budeme používat, není příliš složité, tak určení (odhadu) variability našich dat či velikosti následně naměřených efektů snadné naopak není. U studií, které jsou replikacemi se lze opřít o již získané výsledky. U studií, které jsou novátorské, pak lze využít buď výzkumnickou intuici, nebo různá doporučení založená na tzv. mírách věcné významnosti (viz dále).

Klíčové shrnutí

Pro stanovení velikosti výzkumného souboru z pohledu statistiky potřebujeme znát:

- 1) Očekávanou variabilitu našich dat
- 2) Požadovanou šíři intervalu spolehlivosti
- 3) Statistickou techniku, kterou využijeme pro zpracování
- 4) Odhad velikosti zkoumaného efektu





3 STANOVENÍ VELIKOSTI VÝZKUMNÉHO SOUBORU DLE POŽADOVANÉ CHYBY (PŘÍSTUP ZALOŽENÝ NA ŠÍŘI INTERVALŮ SPOLEHLIVOSTI).

První přístup, se kterým se krátce seznámíme je přístup založený na požadované šířce intervalu spolehlivosti. Jde o přístup výpočetně velmi jednoduchý, v praxi sociálněvědního výzkumu však velmi málo užívaný. Hlavním důvodem je skutečnost, že umožňuje stanovení velikosti souboru jen pro jednodušší statistické postupy: odhad střední hodnoty či rozdílů dvou středních hodnot² a odhad proporce či rozdílů dvou proporcí³. Další důvod pro malé využívání tohoto postupu je skutečnost, že v sociálněvědních šetřeních většinou zkoumáme mnoho charakteristik najednou⁴ a pro každou z nich bychom museli provést výpočet a výslednou velikost souboru bychom pak stanovili jako maximum těchto vypočtených hodnot.

Pro první vhléd je vhodné se s těmito přístupy seznámit a pochopit logiku jejich využití. Pro ilustraci zvolíme dva nejjednodušší postupy: postup pro odhad střední hodnoty a postup pro odhad proporce.

3.1 Postup pro odhad střední hodnoty

Začneme situací, kdy nás zajímá odhad střední hodnoty některé z kardinálních proměnných. Intervalový odhad střední hodnoty je dán vzorcem 1:

$$P(\bar{x} - t_{1-\alpha/2} \cdot \frac{s'_x}{\sqrt{n}} < \mu < \bar{x} + t_{1-\alpha/2} \cdot \frac{s'_x}{\sqrt{n}}) = 1 - \alpha \quad (1),$$

kde $1-\alpha$ je požadovaná spolehlivost (obvykle 95%)

n – velikost výzkumného souboru

$t_{1-\alpha/2}$ - $1-\alpha/2$ procentní kvantil t-rozdělení s $n-1$ stupni volnosti

² Tedy o situaci, kterou zpracováváme skrze t-testy.

³ Tedy situace, kterou zpracováváme skrze binomické testy.

⁴ Výjimkou nejsou ani stovky proměnných v jednom výzkumu.





μ - odhadovaná střední hodnota

s_x – směrodatná odchylka naší proměnné

\bar{x} - - průměr naší proměnné

Z tohoto vzorce se dá poměrně snadno odvodit vzorec 2, který ukazuje jak lze pro danou spolehlivost, směrodatnou odchylku a požadovanou šíři intervalu spolehlivosti stanovit minimální velikost souboru.

$$n \geq \frac{t_{1-\alpha/2}^2 s_x'^2}{\Delta^2} \quad (2), \text{ kde}$$

Δ – požadovaná polovina šíře intervalu spolehlivosti

ostatní symboly mají stejný význam jako ve vzorci 1

Pro větší soubory (řádkově nad 30 jednotek) lze užít místo kvantilů t-rozdělení kvantily normálního rozdělení, tj. ve vzorci 2 bude místo $t_{1-\alpha/2}$ $u_{1-\alpha/2}$. Zkusme si ukázat dva příklady výpočtu založené na vzorci 2.

Příklad 1.

Chceme zjistit, jak velký soubor bude potřebný pro odhad střední hodnoty příjmů. Předpokládáme, že variabilita našich dat vyjádřená směrodatnou odchylkou je 10000 Kč. Požadujeme, aby výsledný 95% interval spolehlivosti měl šířku:

a) ± 2000

b) ± 1000

Aplikace vzorce 2 na naše zadání je uvedena v příloženém souboru v Excelu.

Výsledné řešení pro situace a) a b) je následující:

a) 384 jednotek

b) 1537 jednotek.





Pokud tedy budeme chtít 95% interval spolehlivosti se šíří ± 2000 Kč, budeme potřebovat 384 jednotek, při požadavku na poloviční šíři (± 1000), bude třeba zjistit informace od 1537 jednotek. Empicky tedy potvrzujeme dříve formulované pravidlo, že pro 2x větší přesnost odhadu je třeba 4x větší soubor.⁵

3.2 Postup pro odhad proporce

Zřejmě ještě častěji, než odhad střední hodnoty realizujeme v sociálních vědách odhad proporce. Tato skutečnost plyne z toho, že většina našich proměnných nemá kardinální charakter a zajímá nás typicky (relativní) četnost výskytu jediné kategorie naší proměnné. V případě intervalového odhadu se vzorec 1 ještě zjednodušuje, protože v něm přímo nefiguruje variabilita proměnné (jak již víme je odvozena od úrovně příslušného jevu, tj. od příslušné proporce). Interval spolehlivosti pro proporce udává vzorec 3.

$$P\left(p - u_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} < \pi < p + u_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha \quad (3), \text{ kde}$$

$1-\alpha$ je požadovaná spolehlivost (obvykle 95%)

n – velikost výzkumného souboru

$u_{1-\alpha/2}$ - $1-\alpha/2$ procentní kvantil normálního rozdělení

p – proporce výskytu v našich datech

π – proporce v populaci

Z tohoto vzorce lze opět poměrně snadno odvodit vzorec 4, který lze využít pro výpočet minimální požadované velikosti souboru:

$$n \geq \frac{u_{1-\alpha/2}^2 \pi(1-\pi)}{\Delta^2} \quad (4), \text{ kde}$$

Δ – požadovaná polovina šíře intervalu spolehlivosti

ostatní symboly mají stejný význam jako ve vzorci 3

⁵ Matematicky tato vlastnost plyne ze vzorce 2, kde n závisí nepřímo úměrně na druhé mocnině Δ .





Aplikaci vzorce 4 opět demonstrujeme na příkladu.

Příklad 2.

Chceme zjistit jak velký soubor bude potřebný pro zjištění proporce voličů strany A. Předpokládáme, že strana A má podporu 20 % voličů. Požadujeme, aby výsledný 95% interval spolehlivosti měl šířku:

a) ± 4 p.b.⁶

b) ± 2 p.b.

Výpočet dle vzorce 4 pro naše zadání je opět uveden v příloženém souboru v Excelu.

Výsledné řešení pro situace a) a b) je následující:

a) 384 jednotek

b) 1587 jednotek

Pokud tedy chceme 95% interval spolehlivosti se šíří ± 4 procentní body, budeme potřebovat 384 jednotek, při požadavku na poloviční šíři (± 2 p.b.), bude třeba zjistit informace od 1537 jednotek. Díky vhodnému zadání jsou výsledky analogické příkladu 1.

Příklad 3.

Ještě krátce navážeme na příklad 2 a ukážeme, jak by se měnil požadavek na velikost souboru, pokud by strana A měla menší nebo větší proporce příznivců. Budeme uvažovat dvě varianty:

a) 10% podporu

b) 50% podporu

Výpočty jsou opět v Excelu.

a) 216 jednotek

⁶ P.b. – procentní body, tj. rozdíl mezi dvěma údaji v procentech.





b) 600 jednotek

Z výpočtů je tedy patrné, že pro jevy, jejichž proporce je malá (nebo extrémně velká) je požadavek na výzkumný soubor menší, než pro jevy, které se proporcí pohybují okolo 50 %, jak bylo naznačeno již na počátku textu.



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



4 PŘÍSTUP ZALOŽENÝ NA MINIMÁLNÍ POŽADOVANÉ SÍLE TESTU

Po rozebrání přístupů založených na šířce intervalu spolehlivosti se budeme věnovat v praxi daleko častěji užívanému přístupu, který je založen na požadované síle testu. Pro detailnější vhléd musíme provést definici pojmu síla testu⁷. Formálně jde o podmíněnou pravděpodobnost, že zamítneme nulovou hypotézu, pokud tato hypotéza neplatí. Srozumitelněji vyjádřeno jde tedy o schopnost statistického testu detekovat efekt. Ještě jiný způsob vysvětlení by mohl mít empirický charakter. Předpokládejme opět příklad se dvěma skupinami (klasická výuka matematiky a Hejného metoda) a předpokládejme, že existuje rozdíl těchto dvou přístupů změřený na celé populaci žáků (například o velikosti 5 bodů v testu, který má směrodatnou odchylku 10 bodů). Hypoteticky bychom mohli provádět opakované náhodné výběry žáků (např. 1000 výběrů) a testovat, zda existuje rozdíl středních hodnot. Proporce testů, které by indikovaly rozdíl by byla právě síla testu.

Intuitivně tušíme, že síla statistického testu by měla být poměrně vysoká, jinak bychom používali vědecké postupy zbytečně. Existují mnohá doporučení ohledně požadavku ve minimální síle testu. Nejslavnější je doporučení Cohena (1988), který ve své knize opakovaně apeluje, aby byla síla testu minimálně 80 % (tj. pravděpodobnost minimálně 0,8). Dnes se poměrně často setkáváme i s vyššími hodnotami (např, v oblasti medicínského výzkumu není výjimkou požadavek na 90% či 95 % sílu).

Výpočty velikosti souboru lze provádět pro jednodušší situace za pomoci kalkulačky (vzorce nabízí ve své knize již zmíněný Cohen, 1988; velice zjednodušené postupy pak kniha Kraemer&Blasey, 2016), nicméně v praxi převažuje výpočet v software. Programů je celá řada, pro ilustrace výpočtů je dále použit freeware G*Power, který je jeden z nejpoužívanějších. Před praktickými ukázkami ještě připomeneme, co bude třeba pro výpočet požadované velikosti výzkumného souboru zadat. Budou to tyto parametry:

- a) statistickou techniku (statistický test), kterou využijeme pro zpracování,

⁷ Symbolické značení síly testu je obvykle $1-\beta$.





- b) požadovanou hladinu statistické významnosti (alfa)⁸
- c) požadovanou minimální velikost síly testu a
- d) odhad velikosti zkoumaného efektu.

Na základě těchto tří nastavení pak software spočítá požadovanou velikost souboru. První ilustraci praktického výpočtu budeme směřovat ke korelacím.

⁸ V praxi často volíme $\alpha=0,05$, postup odpovídá stanovování 95% intervalů spolehlivosti.





4.1 Stanovení velikosti souboru pro korelace

Příklad 4.

Chceme zjistit, jak velký soubor bude potřebný pro zjištění korelace mezi dvěma veličinami. Předpokládáme, že síla testu by měla být minimálně 80 %, hladina statistické významnosti 5 %, budeme využívat Pearsonův korelační koeficient a námi očekávaná korelace veličin je 0,3.

Spustíme program G*Power a provedeme zadání dle obrázku 1.

Obrázek 1. Zadání výpočtu velikosti souboru pro korelační studii (Pearson, oboustranný test pro $r=0,3$, $\alpha=0,05$, $1-\beta=0,8$)

The screenshot shows the G*Power software interface with the following settings:

- Test family:** Exact
- Statistical test:** Correlation: Bivariate normal model
- Type of power analysis:** A priori: Compute required sample size - given α , power, and effect size
- Input Parameters:**
 - Tail(s): Two
 - Determine => Correlation ρ H1: 0.3
 - α err prob: 0.05
 - Power ($1-\beta$ err prob): 0.8
 - Correlation ρ H0: 0
- Output Parameters:**
 - Lower critical r: ?
 - Upper critical r: ?
 - Total sample size: ?
 - Actual power: ?

Buttons at the bottom: Options, X-Y plot for a range of values, Calculate.

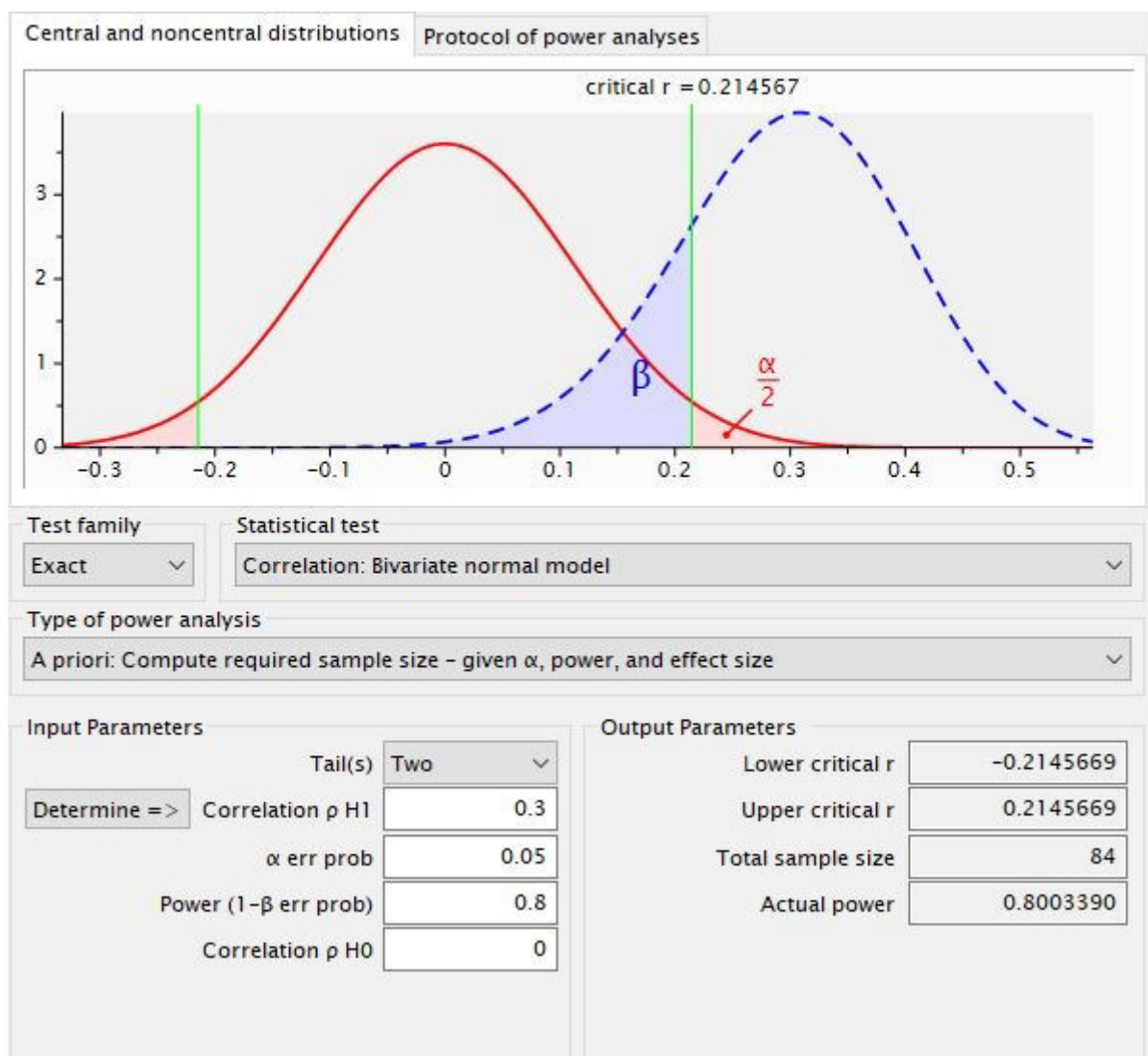
Nejdříve musíme vybrat statistickou techniku (tj. Pearsonův korelační koeficient). V levém horním rolovníku *Test family* vybereme *Exact* a ve vedlejším rolovníku *Statistical Test* pak zvolíme *Correlation: Bivariate normal model*. Poté v poli *Input Parameters* zadáme do políčka *Tails* rolovníkem variantu *Two* (oboustranný test), do pole *Correlation H1* hodnotu 0,3, do políčka *α err prob* hodnotu 0,05 a do políčka *Power* hodnotu 0,8. Takto je připra-





ven výpočet a stačí v pravém dolním rohu stisknout tlačítko *Calculate*. Po jeho spuštění dostaneme výsledek, který ukazuje obrázek 2.

Obrázek 2. Výsledek výpočtu velikosti souboru pro korelační studii (Pearson, oboustranný test pro $r=0,3$, $\alpha=0,05$, $1-\beta=0,8$)



Výpočetní výsledek najdeme vpravo vedle zadání, doprovodné grafické výsledky jsou v horním velkém rámečku. Nadto lze ještě vyžádat další grafické výstupy (nebudeme je zde prezentovat). Základní výsledek (požadovaná velikost souboru) je v políčku *Total*





sample size. Pro naše zadané parametry bychom tedy potřebovali minimálně 84 výzkumných jednotek.

Horní graf ukazuje dvě křivky – červená ukazuje rozdělení testování statistiky při platnosti nulové hypotézy (nulové korelace), modrá pak při platnosti alternativní hypotézy (korelace o velikosti 0,3). Červená plocha pak ukazuje polovinu hladiny statistické významnosti ($\alpha/2^9$), modrá plocha pak doplněk velikosti síly testu do jedničky (β), tj. v našem případě hodnotu 0,2.

Pro možné úvahy lze zmínit několik souvislostí:

1. Pokud by námi očekávaná korelace byla nižší, požadavek na velikost souboru by *ceteris paribus* vzrostl (např. pro $r=0,1$ bychom potřebovali 782 jednotek). Naopak pro větší očekávanou korelaci by stačil i menší soubor (např. pro $r=0,5$ bychom potřebovali 29 jednotek).
2. Pokud bychom vyžadovali vyšší sílu testu, potřebovali bychom *ceteris paribus* větší soubor. Například pro sílu 90 % by pro korelaci o velikosti 0,3 bylo zapotřebí 112 jednotek, pro 95 % sílu pak 138 jednotek.
3. Pokud bychom vyžadovali nižší hladinu statistické významnosti, potřebovali bychom *ceteris paribus* také větší soubor. Například pro $\alpha=0,01$ by pro korelaci o velikosti 0,3 bylo zapotřebí 125 jednotek, pro $\alpha=0,005$ pak 142 jednotek.

Na okraj je třeba ještě poznamenat, že pokud bychom přešli od Pearsonovy korelace pro kardinální data ke korelacím pro ordinální proměnné (běžně užívané jsou koeficienty dle Spearmana a Kendalla) požadavek na velikost souboru opět mírně vzroste.¹⁰

4.2 Stanovení velikosti souboru pro rozdíl středních hodnot (t-testy)

Druhou situací, kterou budeme krátce demonstrovat bude situace pro srovnání středních hodnot dvou nezávislých skupin, nejčastěji zde užíváme dvouvýběrový t-test. Před přecho-

⁹ Pro naše zadání je tedy velikost této plochy 2,5 %.

¹⁰ G*Power ale výpočty pro tyto situace neumí.





dem k praktickým ukázkám bude třeba ještě drobné vysvětlení. Týká se ukazatele, kterému říkáme Cohenovo d . V případě že srovnáváme dvě skupiny z hlediska jejich středních hodnot, bývá zvykem tento rozdíl vyjádřit ve standardizované podobě, tj. po odstranění měřicí stupnice. Právě tuto operaci zajišťuje Cohenovo d , které je rozdílem středních hodnot mezi dvěmi skupinami, kdy tento rozdíl vyjadřuje relativně vůči velikosti směrodatné odchylky našich dat. Formálně tedy vypočteme Cohenovo d dle vzorce 5.

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}, \quad (5) \text{ kde}$$

d – Cohenovo d

\bar{x}_1 - je průměr první skupiny

\bar{x}_2 - je průměr druhé skupiny a

s – směrodatná odchylka našich dat (obou skupin dohromady).

Vrátíme-li se k našemu výkladu s Hejného metodou, pak při rozdílu 5 bodů v testu při směrodatné odchylce 10 je velikost Cohenova $d=0,5$ (5/10). Cohen ve své knize (1988) mj. stanovil doporučené interpretace hodnot ukazatele d , doporučení reprodukuje tabulka 1.

Tabulka 1. Velikosti Cohenova d a jejich doporučená interpretace

Cohenovo d	0,2	0,5	0,8
Označení	Malý rozdíl	Střední rozdíl	Velký rozdíl

Zdroj: Cohen, 1988

Zdůrazněme, že Cohenovská doporučení je nutno brát jen jako orientační, moou nám ale dobře posloužit pro výpočty požadované velikosti souboru. Budeme tedy tyto výpočty provádět variantně pro různé velikosti rozdílů měřených Cohenovým d , abychom měli možnost vidět dopady velikosti zkoumaného efektu.





Příklad 5.

Chceme zjistit, jak velký soubor bude potřebný pro zjištění rozdílu mezi dvěma skupinami. Předpokládáme, že síla testu by měla být minimálně 80 %, hladina statistické významnosti 5 %, budeme využívat dvouvýběrový t-test a námi očekávaný rozdíl mezi skupinami vyjádřeným Cohenovým d bude:

a) $d=0,2$

b) $d=0,5$

c) $d=0,8$

Pro jednoduchost předpokládáme, že obě skupiny jsou stejně velké.

Spustíme opět program G*Power a provedeme zadání dle obrázku 3.





Obrázek 3. Zadání výpočtu velikosti souboru pro dvouvýběrový t-test (oboustranný test pro $d=0,1$, $\alpha=0,05$, $1-\beta=0,8$)

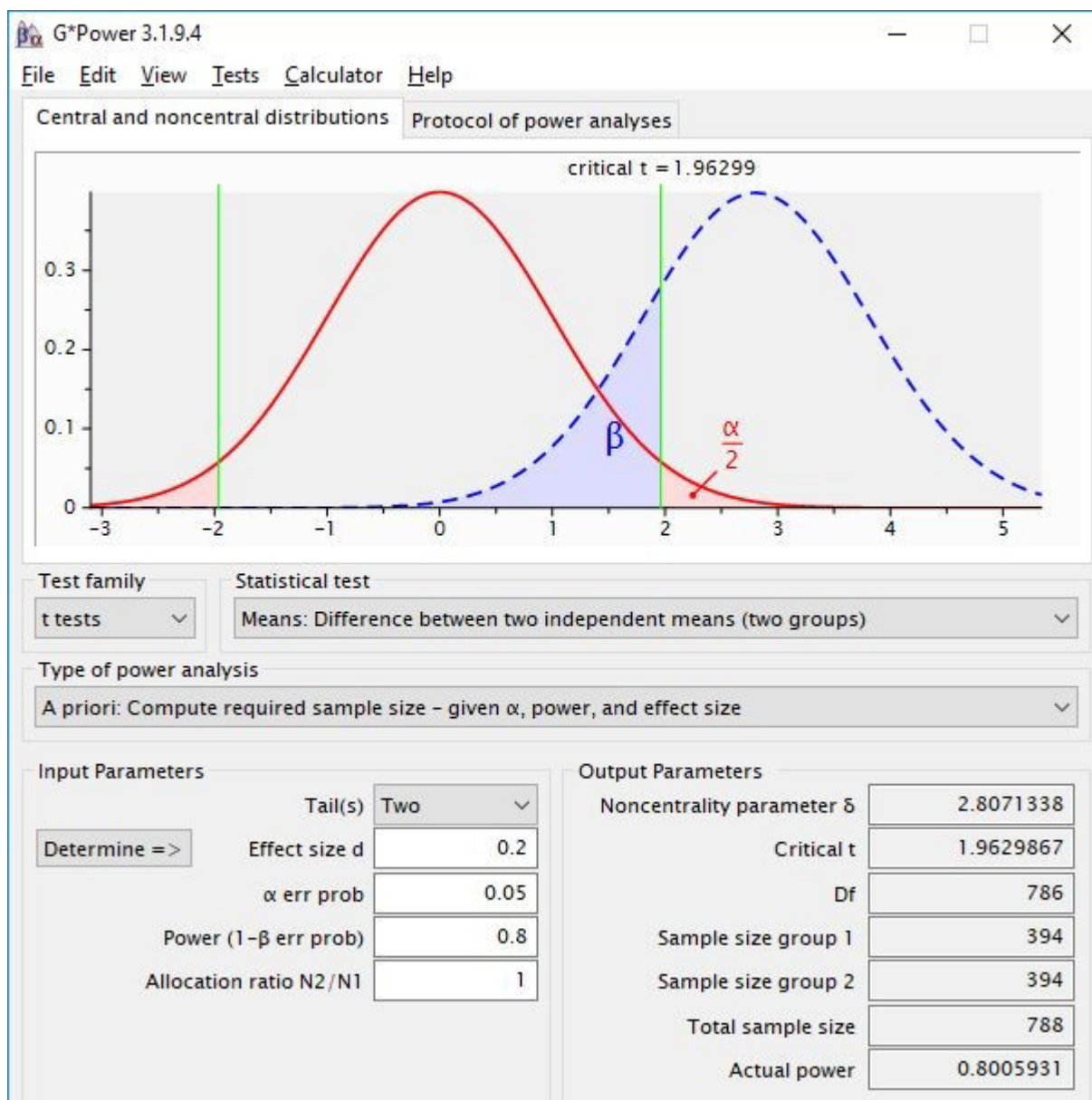
The screenshot shows a software interface for calculating sample size. It is divided into several sections:

- Test family:** A dropdown menu set to "t tests".
- Statistical test:** A dropdown menu set to "Means: Difference between two independent means (two groups)".
- Type of power analysis:** A dropdown menu set to "A priori: Compute required sample size - given α , power, and effect size".
- Input Parameters:** A section with a "Determine =>" button and several input fields:
 - Tail(s): A dropdown menu set to "Two".
 - Effect size d: A text input field containing "0.2".
 - α err prob: A text input field containing "0.05".
 - Power ($1-\beta$ err prob): A text input field containing "0.8".
 - Allocation ratio N2/N1: A text input field containing "1".
- Output Parameters:** A section with several output fields, each with a question mark in a box:
 - Noncentrality parameter δ
 - Critical t
 - Df
 - Sample size group 1
 - Sample size group 2
 - Total sample size
 - Actual power
- Buttons:** At the bottom, there is a grey button labeled "X-Y plot for a range of values" and a blue button labeled "Calculate".

Nejdříve musíme vybrat statistickou techniku (tj. dvouvýběrový t-test). V levém horním rolovníku *Test family* vybereme *t-tests* a ve vedlejším rolovníku *Statistical Test* pak zvolíme *Means: Difference between two independent means*. Poté v poli *Input Parameters* zadáme do políčka *Tails* rolovníkem variantu *Two* (oboustranný test), do pole *Effect size d* hodnotu *0,2*, do políčka *α err prob* hodnotu *0,05* a do políčka *Power* hodnotu *0,8*. S ohledem na stejnou velikost skupin necháme *Allocation ratio* na hodnotě 1. Je připraven výpočet a stačí v pravém dolním rohu stisknout tlačítko *Calculate*. Po jeho spuštění dostaneme výsledek, který ukazuje obrázek 4.



Obrázek 4. Výsledek výpočtu velikosti souboru pro dvouvýběrový t-test (oboustranný test pro $d=0,1$, $\alpha=0,05$, $1-\beta=0,8$)



Výpočetní výsledek najdeme opět vpravo vedle zadání, doprovodné grafické výsledky jsou opět v horním velkém rámečku (ponecháme je bez komentáře). Základní výsledek (požadovaná velikost souboru) je v políčkách *Sample size group 1* a *Sample size group 2*. Pro naše zadané parametry bychom tedy potřebovali minimálně 394 jednotek v každé skupině, celkem (Total sample size) pak 788 jednotek. Nyní lehce odbočíme od technického komentáře a dovolíme si komentář věcný. Pokud by rozdíl středních hodnot dvou skupin byl



0,1 směrodatné odchylky, je třeba mít k dispozici poměrně velký soubor (téměř 800 jednotek). Je tedy namísto si položit otázku, zda takto malé rozdíly mají být vědci zkoumány, protože náklady na sběr dat mohou být u této velikosti souboru enormní,

Vrátíme-li se k zadání příkladu 5, pak zbývá zodpovědět otázku, jak velký by měl být soubor, pokud bychom žádali rozdíl středních hodnot o velikosti Cohenova d 0,5 (b)), resp. 0,8 (c)). Analogickým zadáním jako v obrázku 3 bychom získali tyto výsledky:

Ad b) $N=126$ jednotek (64 v každé skupině) a

Ad c) $N=52$ (26 jednotek v každé skupině).

Z výsledků se opět potvrzuje, že s tím, jak roste velikost námi sledovaného efektu, klesá požadavek na minimální velikost výzkumného souboru. Samozřejmě platí i další závěry, které byly uvedeny u korelací.





SHRNUTÍ

Text předložil základní postupy pro stanovení velikosti výzkumného souboru a ukázal praktické výpočty v MS Excel a G*Power.

Pro stanovení velikosti výzkumného souboru na základě šířky intervalu spolehlivosti potřebujeme znát:

- 1) variabilitu našich dat,
- 2) požadovanou šíři intervalu spolehlivosti,
- 3) statistickou techniku, kterou využijeme pro zpracování a
- 4) odhad velikosti zkoumaného efektu.

Pro stanovení velikosti výzkumného souboru na základě síly testu potřebujeme znát:

- 1) požadovanou sílu testu,
- 2) požadovanou hladinu statistické významnosti,
- 3) statistickou techniku, kterou využijeme pro zpracování a
- 4) odhad velikosti zkoumaného efektu





SEZNAM POUŽITÉ LITERATURY

Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Lawrence Erlbaum Associates, Publishers.

Kraemer H. Ch., Ch. Blasey. 2016. How Many Subjects?: Statistical Power Analysis in Research (2nd ed.). Sage.



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

**MS
MT**
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



SEZNAM PŘÍLOH

Příloha k výpočtům velikosti: 1xMS Excel soubor



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



Příklad 1

spolehlivost 95%

směr. odchylka 10000

šíře int. spol.:

a) 1000

b) 500

Výpočet

a) 384,1459

b) 1536,584

poměr b:a 4

$$n \geq \frac{t_{1-\alpha}^2 S_x'^2}{\Delta^2}$$

Příklad 2

spolehlivost	95%
proporce	0,2
šíře int. spol.:	
a)	0,04
b)	0,02

Výpočet

a)	384,1459
b)	1536,584

poměr b:a 4

Příklad 3

a) proporce	0,1	216,0821
b) proporce	0,5	600,2279

$$n \geq \frac{u_{1-\alpha/2}^2 \pi (1 - \pi)}{\Delta^2}$$