



Může Bayesův faktor nahradit P hodnotu?

Is Bayes Factor Capable of Replacing P Value?

Petr Soukup

ABSTRACT The P value was introduced as a value to evaluate the results of statistical tests. The basic concept originated in the 1920s, and after the Second World War it was significantly expanded. For about the last three decades there has been intense discussion about the problematic features of the P value concept and its use in science, and voices calling to abolish use of the P value are growing louder. In addition, suggestions have been made for alternative procedures that could replace or supplement the P value. Statisticians have tried to invent an indicator similar to the P value, but without its weaknesses. There are many of these options. Besides alternatives within the classical statistical testing paradigm, the use of an alternative statistical approach, so-called Bayesian statistics, is increasingly being discussed. An example of a moderate recommendation is that of using the Bayes factor, essentially an analogue of the P value in the Bayesian world. The aim of this article is to present the Bayes factor in detail, to describe its similarities and dissimilarities with the P value, and discuss the possibilities of its calculation. In addition to computational procedures, a detailed discussion of the weaknesses of the Bayes factor is also included.

KEYWORDS quantitative data analysis, statistical testing, P value, Bayes factor

V oblasti sociálněvědního kvantitativního zpracování dat dominuje užívání P hodnoty, resp. statistických testů. Tuto koncepci přinesl světu ve 20. letech 20. století zejména Fisher, na kterého navázali Neyman s Pearsonem. Především po druhé světové válce se začalo užívání P hodnoty ve vědě rozšiřovat a díky příchodu výpočetní techniky se stalo téměř automatickým postupem. Nutno ale dodat, že zvláště v posledních desetiletích sílí kritika P hodnoty. Ostatně tato kritika je zhruba stejně stará jako koncepce sama. Mnohé argumenty lze nalézt v článku Soukupa (2010) a literatuře tam uvedené. Existují i empirické důkazy problémového užívání, v českých poměrech zejména v textech Cubereka s Frömelem (2011) a Soukupa (2019), ze zahraničí z nedávné doby například článek kolektivu vedeného Bernardim (2017). Není proto vůbec překvapivé, že v oficiálních doporučeních odborných asociací APA (2010) či AERA (2006) se objevují varování před přeceňováním role P hodnoty a je vyžadováno užívání alternativ (například intervalů spolehlivosti) či jejich doplňků (zejména měř věcné významnosti, více lze nalézt v textu Soukupa [2013] a literatuře tam

uvedené). Některá doporučení však jdou ještě dále. Z těch nejznámějších si zaslouží zmínku stanovisko ASA (Wassserstein a Lazar 2016), které mj. doporučuje užívání bayesovské statistiky či jejích nástrojů. Na toto stanovisko ještě navázalo první číslo časopisu *American Statistical Association* v roce 2019, které obsahovalo speciální přílohu čítající 43 článků předních statistiků vyslovujících se k problémům spojeným s P hodnotou a jejími alternativami.¹ Právě Bayesův faktor lze považovat za jednu z nejjednodušších alternativ k P hodnotě a díky vývoji softwaru také čím dál použitelnější (existují online kalkulačky, pomůcky v Excelu, některé procedury v běžně užívaném SPSS a mnohé speciálnější postupy v R či bayesovském softwaru, jako jsou WinBugs, Stan či Jags).

Definice P hodnoty a její problémy

Před samotným představením Bayesova faktoru je vhodné připomenout definici P hodnoty (často v softwaru označované jako P-value, Sig či alfa) a také připomenout stručně její základní nedostatky. P hodnota typicky slouží pro vyhodnocení našich výzkumných hypotéz a ideální podklad pro rozhodování by například zněl, že naše zkoumaná hypotéza (často formulovaná jako alternativní) má pravděpodobnost 0,8 (či 80%), naopak k ní doplňková nulová hypotéza pak 0,2 (resp. 20%). Ovšem toto (pro nás potřebné) sdělení P hodnota rozhodně neumí (předběhneme a sdělíme čtenáři, že to neumí ani Bayesův faktor). Co nám tedy P hodnota říká? Je to pravděpodobnost, že při platnosti nulové hypotézy (typicky hovoříci o nulovém rozdílu či nulové souvislosti) získáme naše data, nebo data ještě extrémnější (míněno např. s vyšším průměrem či jinou zkoumanou charakteristikou). Zde se objevuje první námitka proti P hodnotě: jde o hodnotu, která je zčásti založena na datech, která jsme nezískali (to jsou ty výsledky „extrémnější než námi získané“ v definici P hodnoty).

Další problém je, že P hodnota primárně hovoří o datech, a ne hypotézách, ostatně formálně matematicky jde o podmíněnou pravděpodobnost $P(D/H_0)$. A pokud o nějaké hypotéze P hodnota hovoří, pak je to maximálně hypotéza nulová (H_0), nikoli alternativní (H_1). Přitom většina analýz končí tím, že zamítáme nulovou hypotézu a rádi bychom vyčíslili pravděpodobnost přijaté alternativní hypotézy při existenci našich dat, tj. formálně $P(H_1/D)$. Dalším běžně známým problémem P hodnoty je závislost na velikosti výběrového souboru. Platí totiž, že ceteris paribus (při stejné velikosti zkoumaného efektu) se P hodnota se zvětšujícím výběrem snižuje. Trochu s nadsázkou lze říci, že aby byl výsledek statistiky průkazný (významný), stačí mít jen dostatečně velký datový soubor. To je ale jistě nevhodná logika pro vědeckou práci. P hodnotě bývá vytýkáno též (mírně neprávem), že podporuje dichotomické rozhodování, zamítání nebo nezamítání nulové hypotézy. Co stojí za pozornost, je skutečnost, že P hodnota se chová macešsky k nulové hypotéze, zatímco u alternativní umožňuje její přijetí, u nulové nikoli.²

¹ Detailní rozbor této přílohy (více než 400 stran) by vyžadoval minimálně plochu jednoho článku. Je smutné, že reflexe těchto textů jak na mezinárodní, tak národní úrovni je zatím mizivá, světlou výjimkou jsou některé texty v metodologicky orientovaných časopisech pro psychology.

² Pomíjíme skutečnost, že někteří autoři a mnozí studenti ve svých textech hovoří o přijetí nulové hypotézy. To není chyba P hodnoty, ale konkrétních autorů těchto textů. P hodnota takový závěr neumožňuje.

Zde je nutno upozornit, že existují návrhy (v praxi ovšem přehlížené), které tento nedostatek alespoň částečně překonávají (srov. např. 4 pásma pro P hodnotu v textu Wasserman 2004). Další problémy P hodnoty a situace, kdy její užití není vhodné či snadné, lze nalézt v článku Soukupa a Rabušice (2007). Ponecháváme bez povšimnutí, že definice P hodnoty je poměrně komplikovaná a většina jejích uživatelů jí nerozumí, a proto ji užívá špatně. Klasický test Oakese (1986) byl několikrát opakovaně využit (Haller a Krauss 2002) a všechny jeho známé replikace (Gigerenzer 2018) ukazují, že koncepci P hodnoty nerozumí nejen studenti, ale ani akademici (bohužel včetně těch, kteří vyučují sociálněvědní statistiku). V Česku zatím studii na toto téma nemáme, autor tohoto článku ale již více než 10 let testuje pomocí Oakesova nástroje (1986) a výsledky odpovídají zahraničním zkušenostem.

Motivační příklad k Bayesovu faktoru

Před zavedením formální definice Bayesova faktoru a ukázkou „ručního“ výpočtu³ bude uveden jednoduchý motivační příklad.⁴ Mějme dva výzkumníky, zástupce dvou výzkumných agentur, pracovně je označme jako A a B. Tito dva výzkumníci mají velice precizní představu o voličské podpoře strany XY. Konkrétně výzkumník A tvrdí, že tato strana má podporu 22 % občanů, a výzkumník B tvrdí, že strana má podporu 26 % občanů. Zde narážíme na první rys bayesovské statistiky, tj. již před získáním dat a výpočty z nich musíme formulovat svou představu o tom, co chceme zjistit. Formálně potřebujeme mít představu o neznámém populačním parametru (v našem případě jde o proporcii příznivců strany XY). Tuto představu označujeme ve statistice jako apriorní rozdělení. Může být velice precizní (náš případ, tj. je to přesně 22 %, resp. je to přesně 26 %⁵), nebo naopak zcela vágní⁶ (například je to mezi 0 a 100 %). **Poté co stanovíme apriorní rozdělení, získáme data a pomocí kombinace našich dat (jejich věrohodnostní funkce⁷) a apriorního rozdělení získáme tzv. aposteriorní rozdělení.** Toto je v kostce popis postupu bayesovské analýzy dat. **Bayesův faktor pak v rámci tohoto postupu ukazuje, jaká je změna mezi apriorním a aposteriorním rozdělením, jinak řečeno, jak naše data změnila naše představy, které jsme měli na počátku.** Konkrétně jde pak o poměr toho, jak se změnila představa o dvou

³ Není třeba se bát, že by pro výpočty bayesova faktoru bylo třeba užívat ručních výpočtů. Na vše jsou dnes k dispozici programy či pomůcky (online nebo pro Excel). Ostatně pro většinu situací bayesův faktor ani ručně počítat nelze. To je ostatně určitě jeden z důvodů, proč je užíván poměrně málo, donedávna nebylo možné jej spočítat a ani v současnosti nejde o zcela běžně dostupný výpočet ve statistickém softwaru.

⁴ Příklad je inspirován příkladem na blogu Moreyho (<http://bayesfactor.blogspot.com/2014/02/the-bayesfactor-package-this-blog-is.html>). Jde o jednoho z autorů, který se detailně věnuje problematice bayesova faktoru. Moreyho výklad je mnohem komplexnější, zde uvedený příklad je přípustným zjednodušením pro první vhléd do problematiky.

⁵ Zde jde samozřejmě o výrazné zjednodušení. Používá se pro snadný výklad, v praxi takto precizní představy nikdo formulovat nebude.

⁶ Výraz vágní (*vague*) zde užívají i statistici, alternativně se užívá výrazů neinformativní (*noninformative*) rozdělení či difúzní (*diffuse*) rozdělení.

⁷ Jde o funkci, která pro jednotlivé hodnoty parametrů určuje pravděpodobnost získání našich dat.

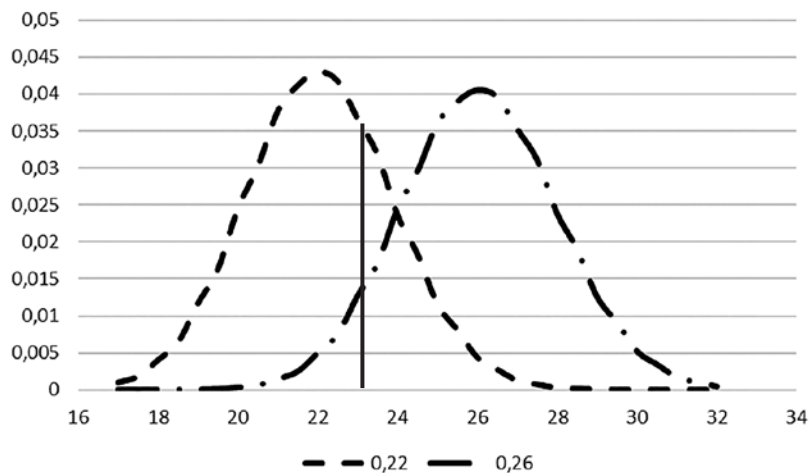
konkurujících hypotézách (typicky nulové a alternativní). Vraťme se k našemu příkladu. V našem příkladu nemáme nulovou a alternativní hypotézu, ale máme hypotézu výzkumníka A a výzkumníka B. Poměr jejich změn ve světle získaných výzkumných dat bude Bayesův faktor, který nám umožní přiklonit se spíše k výzkumníkovi A, nebo B. Dejme tomu, že náš sběr dat proběhl na souboru 500 respondentů a z těchto 500 respondentů vyslovilo 23 % podporu straně XY. Laicky usoudíme, že 23 % je blíže hypotéze výzkumníka A, a že tedy jeho hypotéze budeme věřit více. Zapojíme-li jednoduchý statistický aparát, **můžeme vypočítat pravděpodobnost podpory ve výši 23 % při platnosti hypotéz výzkumníka A a B. Formálně půjde o určení pravděpodobnosti pro binomické rozdělení⁸ s parametry (500;0,22), resp. (500;0,26) v bodě 115.⁹** Výpočet lze provést např. v Excelu či statistickém softwaru, konkrétní hodnoty těchto pravděpodobností jsou: 0,367 pro hypotézu výzkumníka A a 0,127 pro hypotézu výzkumníka B. Bayesův faktor je podíl těchto dvou pravděpodobností, tj. $BF_{A/B} = 0,367/0,127 = 2,9$. Jinými slovy, hypotéza výzkumníka A je ve světle našich dat cca 3krát pravděpodobnější než hypotéza výzkumníka B. Pokud bychom se tedy po získání našich dat rozhodovali, komu více věřit, pak by to byl nejspíše výzkumník A. Dodejme, že 3krát vyšší pravděpodobnost ale není nijak převratná, možná bychom tedy měli být ve svém závěru opatrnější a nefavorizovat do získání dalších dat nikoho (ostatně v další části věnované praktickým ukázkám si řekneme, že hodnoty BF do velikosti cca 3 nemáme za žádný jasný důkaz považovat). Nahlédněme ještě na náš příklad graficky (ostatně to bude možné posléze provést i pro prakticky řešené reálné příklady). Zobrazíme si pravděpodobnosti hypotéz výzkumníka A a B pro různé výsledky (Obrázek 1), které bychom mohli získat z výzkumu (omezíme se na volební zisk v intervalu mezi 16 % a 34 %). Bayesův faktor z Obrázku 1 získáme velmi snadno.

Najdeme výšku zobrazené hustoty v bodě 23 pro levou křivku (výzkumník A) a pro pravou (výzkumník B) a tyto dvě hodnoty podělíme (do Obrázku 1 byla pracovně přidána svislá čára, která tyto výšky zobrazuje). Obrázek 1 nám mj. také umožňuje nalézt bod, kde je podpora evidence pro obě hypotézy totožná. Jde o 24% podporu, zde se obě křivky na grafu protínají. Zde by byl Bayesův faktor jedna. Pokud by byla podpora strany XY získaná ve výzkumu pod touto hranicí, bude $BF_{A/B}$ **větší než jedna (tj. ve světle našich dat bude pravděpodobnější hypotéza výzkumníka A). Naopak pokud by byla podpora z výzkumu více než 24%, pak by byl Bayesův faktor $BF_{A/B}$ menší než jedna a data by spíše svědčila ve prospěch hypotézy výzkumníka B.** Shrňme si tuto poučku opět graficky (Obrázek 2).

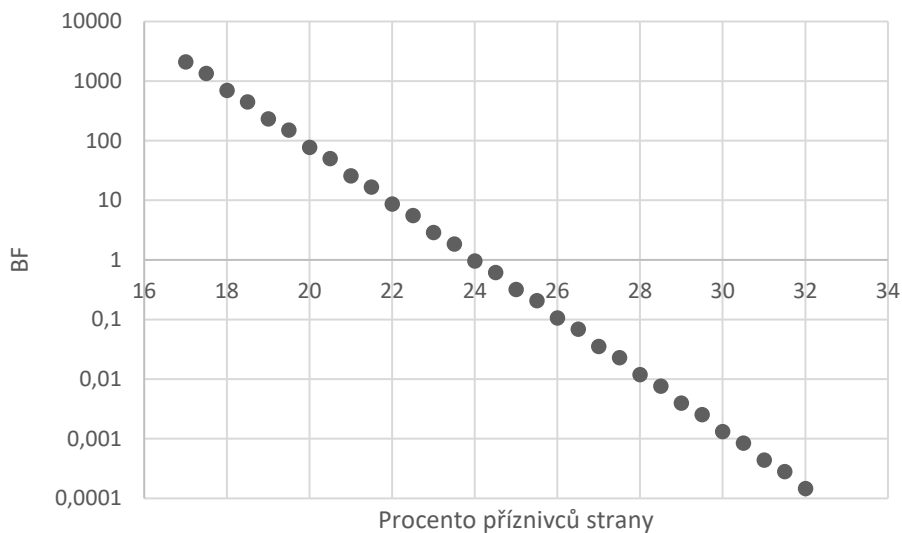
⁸ V případě, že naším zkoumaným fenoménem je proporce, pak příslušné pravděpodobnostní rozdělení je binomické, založené na dvou parametrech – pravděpodobnosti úspěchu a počtu nezávislých pokusů.

⁹ 115 z 500 občanů ve výzkumu vyjádřilo podporu straně XY, to odpovídá 23 %.

Obrázek 1: Hustoty pravděpodobností pro hypotézu výzkumníka A (22% podpora) a B (26% podpora) pro různé míry podpory získané z výzkumu na souboru o velikosti n=500



Obrázek 2: Bayesův faktor pro srovnání hypotézy výzkumníka A (22% podpora) a B (26% podpora) pro různé míry podpory získané z výzkumu na souboru o velikosti n=500



Poznámka: Měřítka na ose Y je logaritmické, využit je dekadický logaritmus o základu 10. Pokud by bylo užito původní měřítka, měla by křivka nelineární průběh.

Nahlédneme-li na Obrázek 2, pak snadno zjistíme, že body protínají vísilou osu v bodě 24, tj. pro tuto míru podpory je Bayesův faktor jednotkový. Pro nižší míru podpory je jeho hodnota nad 1, pro vyšší menší než 1. Z grafu lze například vyčíst, že pokud by výsledek výzkumu hovořil o 18% podpoře strany XY, tak by Bayesův faktor byl blízko hodnoty 1000 (přesným výpočtem bychom získali hodnotu 699¹⁰). Tedy hypotéza B by byla ve srovnání s hypotézou A neudržitelná. Zastavme se ale na chvíli u této argumentace. Tvrdíme, že pokud ve výzkumu založeném na dotazování 500 jednotek zjistíme 18% voličskou podporu, tak je pravděpodobnost hypotézy o 22% podpoře 699krát vyšší než pravděpodobnost hypotézy o 26% podpoře. Nelze ale rozhodně tvrdit, že hypotéza výzkumníka A je správná. Náhledem na Obrázek 1 (či pomocí přesného výpočtu) zjistíme, že pravděpodobnost tohoto výsledku při platnosti hypotézy výzkumníka A je pouhých 0,004! Bayesův faktor je tedy relativním srovnáním pravděpodobnosti dvou hypotéz ve světle našich dat, ale neznamená to automaticky, že favorizovaná hypotéza je správná. To pomocí Bayesova faktoru nezjistíme. Museli bychom využít plných možností bayesovské statistiky a vypočítat aposteriorní rozdělení a z jeho hodnot pak učinit absolutní úsudek o našich hypotézách (o jejich pravděpodobnostech). Příklad zakončíme konstatováním, že formálně bychom mohli srovnávat i více než dvě hypotézy (tedy například tři výzkumníky), ale vždy jen párově (tedy vždy jen dvojice výzkumníků). V drtivé většině aplikací se Bayesův faktor užívá pro srovnání dvou hypotéz. Ke srovnání více než dvou hypotéz se ještě vrátíme v rámci příkladu věnovaného regresní analýze.

Bayesův faktor: stručná historie, definice a základní pojmy

Po uvedení motivačního příkladu a prvního výpočtu Bayesova faktoru je vhodné nahlédnout do historie a zjistit, kdy a kde se koncept Bayesova faktoru zrodil. Jako v mnoha oblastech vědecké činnosti i v případě Bayesova faktoru se setkáváme s nejednoznačným určením autora této koncepce. Převážně bývá za prvního autora koncepce pokládán statistik Jeffreys, který jej detailně rozpracoval a postupně rozvíjel v jednotlivých vydání knihy *Theory of Probability* (1939, 1948, 1961). Intelektuálně se tak dá Bayesův faktor považovat za protíváhu P hodnoty zavedené do statistiky zejména Fisherem.¹¹ Ostatně jak Fisher, tak i Jeffreys působili ve stejné době na univerzitě v Cambridge a jejich korespondence (zejména zpočátku značně útočná a nepřátelská) je známa.¹² To, že se ujala P hodnota,¹³ a ne Bayesův faktor, lze vysvětlit poměrně snadno. Výpočet P hodnoty lze provádět v ruce (Fisher pro tento účel

¹⁰ Čtenář může s výpočty libovolně experimentovat, výpočetní podklady i grafy v Excelu jsou online přílohou článku.

¹¹ Nijak tím nezmenšujeme důležitost koncepce Neymana a Pearsona, která Fisherův koncept výrazně rozvinula a upravila.

¹² Fisher například prohlásil o Jeffreysově knize, že hned na prvních stranách obsahuje chybu a že díky tomu je celá špatně.

¹³ Typicky se užívá podivné smíšení koncepcí od Fishera a od Neymana s Pearsonem, většina uživatelů nemá vůbec tušení o dvojici koncepcí a jejich odlišnostech.

sestavil statistické tabulky, dodnes se objevují v učebnicích), naopak Bayesův faktor s výjimkou nejjednodušších situací ručně spočítat nelze.

Jak upozorňují někteří autoři (např. Good 1979), Bayesův faktor (byť ho takto nepojmenoval) stvořil i Alan Turing, známý matematik, který se proslavil zejména algoritmy pro rozluštění vojenských depeší zasílaných během druhé světové války nepřátelským Německem. Bayesův faktor je tak typickým znovuobjevem (Jeffreys ani Turing o své práci vzájemně nevěděli). Pro soudobou statistiku je příznačné, že navazuje výrazně na koncepci Jeffreysa, a proto se dále Turingovu pojetí věnovat nebudeme.

Pro další postup budeme nyní potřebovat zavést tzv. **Bayesův teorém** (Bayes a Price 1763) a na základě něj formálně popsat Bayesův faktor. Bayesův teorém je **založen na myšlence, že na počátku našich vědeckých úvah formulujeme hypotézy a přiřazujeme jim (apriorní) pravděpodobnosti** (ty buď vycházejí z našich subjektivních přesvědčení,¹⁴ nebo z předchozích výzkumných zjištění). Pro jednoduchost uvažujme o formulaci dvou hypotéz jako v klasické statistice, tj. nulové (H_0) a alternativní (H_1). **Poté získáme výzkumná data (D) a snažíme se vyčíslit (aposteriorní) pravděpodobnost našich hypotéz** po sběru těchto dat. Z předchozího popisu už víme, že poměr změn pravděpodobností (mezi apriorními a aposteriorními) definuje Bayesův faktor. Formální podoba Bayesova teorému může být pomocí námi zavedené symboliky pro nulovou a alternativní hypotézu následující:

$$P(H_0/D) = P(H_0) * P(D/H_0) / P(D) \quad (1)$$

$$P(H_1/D) = P(H_1) * P(D/H_1) / P(D) \quad (2)$$

Vyjádřeno slovně: Aposteriorní pravděpodobnost hypotézy (levá strana) je dána součinem její apriorní pravděpodobnosti a věrohodnosti našich dat, vše je normováno (děleno) tzv. úplnou pravděpodobností zisku našich dat ($P(D)$).¹⁵

Budeme-li chtít vyjádřit poměr aposteriorních pravděpodobností obou hypotéz (zvolíme častěji užívaný poměr mezi alternativní a nulovou hypotézou), získáme tento vzorec:

$$\frac{P(H_1/D)}{P(H_0/D)} = \frac{P(H_1)}{P(H_0)} * \frac{P(D/H_1)}{P(D/H_0)} \quad (3)$$

Poměr aposteriorních pravděpodobností (levá strana) je tedy dán součinem poměru apriorních pravděpodobností (první část pravé strany) a Bayesova faktoru (druhá část pravé strany). Pokud by apriorní pravděpodobnosti obou hypotéz byly shodné, pak by se Bayesův faktor shodoval s poměrem aposteriorních pravděpodobností.¹⁶ S ohledem na znalost základní

¹⁴ Zejména v situaci, kdy fenomén nebyl dříve zkoumán.

¹⁵ Vzorec pro úplnou pravděpodobnost zde neuvádíme.

¹⁶ Občas se této možnosti využívá, pokud nemáme o hypotézách žádnou předchozí empirickou evidenci.

logiky Bayesova faktoru z počátečního ilustračního příkladu rekapitulujeme jeho základní matematické vlastnosti:

1. Jde o hodnotu mezi nulou a nekonečnem.
2. Hodnota jedna tvoří pomyslný střed, tj. vyjadřuje situaci, kdy naše data svědčí ve prospěch obou konkurujících hypotéz se stejnou silou.
3. Hodnoty nad jedna znamenají, že data svědčí spíše ve prospěch alternativní hypotézy (hypotézy v čitateli vzorce 3).
4. Hodnoty pod jedna znamenají, že data svědčí spíše ve prospěch nulové hypotézy (hypotézy ve jmenovateli vzorce 3).

Z důvodů nesymetričnosti stupnice Bayesova faktoru (interval $<0,1>$ a interval $<1,\infty$) se občas místo hodnoty Bayesova faktoru užívá jeho logaritmu (typicky dekadického). Ten má hodnoty od minus nekonečna do plus nekonečna a nula tvoří pomyslný střed (vyrovnaná úroveň evidence dat ve prospěch alternativní i nulové hypotézy).¹⁷ V tomto textu nebudeme dále toto logaritmické vyjádření užívat, nicméně v softwaru se poměrně často objevuje.

Po teoretičtějším popisu je dobré se vrátit k Jeffreysově knize (1961) a poukázat na skutečnost, že Jeffreys dal i první doporučení ohledně interpretací hodnot Bayesova faktoru.¹⁸ Uvedeme původní Jeffreysovo doporučení i velmi známé úpravy provedené Kassem s Rafterym (1995: 777)¹⁹ a Wagenmakersem (Wetzels et al. 2011). Vše shrnuje Tabulka 1.

Snadno nahlédneme, že hodnoty Bayesova faktoru cca do velikosti 3 (inverzní hodnoty od 1/3 do 1) všechny klasifikace považují za zcela zanedbatelné a nehodlají na základě nich vynášet soud o tom, že jedna z hypotéz je favorizována nad druhou (srov. výrazy „anecdotal“, „not worth a bare mention“). Můžeme teď tuto informaci doplnit o návrh zjednodušené práce²⁰ s hodnotou Bayesova faktoru. Pokud bude hodnota BF mezi 1/3 a 3, nelze upřednostnit žádnou hypotézu (tj. naše data svědčí zhruba stejně ve prospěch obou). Pokud bude hodnota BF nad 3, pak upřednostníme alternativní hypotézu, naopak bude-li hodnota pod jednu třetinu, upřednostníme nulovou hypotézu. Srovnáme-li tento (stále zjednodušený přístup) s postupem založeným na P hodnotě, je již nyní zřejmý přínos Bayesova faktoru ve srovnání s P hodnotou.

17 Poučený čtenář jistě správně odhadl, že Bayesův faktor je analogií šancí z analýzy kategoriálních dat a logaritmus BF je analogií logitu užívaného v regresním modelování kategoriálních dat (zejména v logistické regresní analýze).

18 Je namístě již nyní vyslovit varování. Tak jako je výrazně zpochybňována Fisherem doporučovaná hranice 0,05 pro P hodnotu, jsou zpochybňována i doporučení Jeffreysa a jeho následovníků ohledně Bayesova faktoru. Obecně mají tato doporučení pro uživatele spíše negativní dopad a vedou k jejich bezmyšlenkové mechanické aplikaci. Ostatně i toto je jeden z důvodů, proč se pravověrní bayesiánci od Bayesova faktoru výrazně odklánějí a považují ho za koncepci, která bude dobrá maximálně jako náhrada P hodnoty pro uživatele, kteří bayesiánskou statistiku užívat nebudou.

19 Kass s Rafterym (1995) mají v článku i jiné členění, to zde necháme bez povšimnutí. Logika jejich doporučení je odvozena od hodnoty kritéria BIC (srov. detailnější výklad na konci článku), které výrazně propaguje zejména Adrian Raftery. Původní logika doporučení Jeffreysa s tímto nemá nic společného, v době publikace 3. vydání Jeffreysovy knihy (1961) nebylo BIC vůbec známo (jeho vznik se datuje do roku 1978).

20 Vřele tento postup nedoporučujeme, ale jde o základní možnost, proto jej zde uvádíme.

Tabulka 1: Slovní hodnocení velikosti Bayesova faktoru dle Jeffreyse, Kasse s Rafteryem a Wagenmakerse

Hodnota BF	Inverze BF	Označení Jeffreyse (1961)	Označení Kass a Raftery (1995)	Označení Wagenmakers
1–3	1/3–1	Not worth more than a bare mention	Not worth more than a bare mention	Anecdotal
3–10	1/10–1/3	Substantial	Positive	Moderate
10–20	1/20–1/10	Strong	Positive	Strong
20–30	1/30–/20	Strong	Strong	Strong
30–100	1/100–1/30	Very strong	Strong	Very strong
100–150	1/150–1/100	Decisive	Strong	Extreme
>150	<1/150	Decisive	Very strong	Extreme

Pomocí P hodnoty zamítneme či nezamítneme nulovou hypotézu (alternativní v prvním případě přijmeme, ve druhém nikoli). U Bayesova faktoru je možné rozhodnutí trojího druhu:

1. Nedokážeme mezi hypotézami rozhodnout.
2. Upřednostníme alternativní hypotézu.
3. Upřednostníme nulovou hypotézu.

Situace 1 zhruba odpovídá nezamítání nulové hypotézy pomocí P hodnoty, situace 2 jejímu zamítání. Situaci 3 pomocí P hodnoty typicky řešit neumíme (v klasickém testování hypotéz nejsme schopni sdělit, že naše data svědčí ve prospěch nulové hypotézy).

Ještě další rys Bayesova faktoru je poměrně příznivý. Navržená hodnota 3 je relativně vysoká a mnoho výsledků je tak ve světle této hodnoty neprůkazných (srov. příklady v článku a část věnovanou srovnání P hodnot a BF na konci článku). Následování tohoto doporučení tak brání k přečeňování nicotných výsledků, což bohužel hranice 0,05 pro P hodnotu neumí.

Doplňme, že představená **zjednodušená práce s Bayesovým faktorem je problematická**, a pokud jej užíváme, **měli bychom publikovat jeho konkrétní hodnotu a ideálně ji srovnat s hodnotami z podobných studií a ve světle těchto hodnot provést její interpretaci**. Tj. užít relativní přístup namísto absolutního, který vyjadřuje tabulka 1 ve všech verzích tam uvedených.

Vraťme se ještě k Tabulce 1 a povšimněme si, že označení Wagenmakerse (jednoho z nejvýraznějších současných propagátorů Bayesova faktoru a bayesovské statistiky vůbec) je jen velice drobnou modifikací původního Jeffreysova pojetí. Wagenmakers zdůrazňuje, že u Jeffreyse jsou dvě označení („substantial“ a „strong“) významově značně podobná, proto navrhl jejich odlišení výrazy „moderate“ a „strong“. Lze zřejmě souhlasit s tím, že toto odlišení je lepší než původní Jeffreysovo.²¹

²¹ Doplníme, že Wagenmakers je velkým obdivovatelem Jeffreyse a na jeho počest nechal pojmenovat i software, vyvíjený na Amsterdamské univerzitě, který mj. obsahuje výpočty Bayesova faktoru. Jde o software JASP, kde akronym skrývá název Jeffrey's Amazing Statistical Program. Tento software bude využit i pro výpočty v tomto článku.

Krátce se ještě vraťme k našemu ilustračnímu příkladu. Bayesův faktor činil pro naše data 2,9. Ve světle výše uvedených doporučení tedy Bayesův faktor nevykazoval přesvědčivou evidenci ve prospěch hypotézy výzkumníka A. Ostatně v rámci ilustračního příkladu bylo na tuto skutečnost již upozornováno.

Poslední hodnota Bayesova faktoru, která si zaslouží pozornost, je **hodnota 1, resp. hodnoty jí blízké** (tj. vzdálené maximálně v řádu desetin). Jeffreys (1961) hovořil o tom, že v těchto případech situace **nevede k žádnému rozhodnutí** („undecisive“). Je to jedno z možných pojetí, jak na hodnoty BF blízké 1 nahlížet. Alternativní pojetí (preferované i autorem článku) pak říká, že hodnoty blízké jedničce svědčí o tom, že v našem případě **nemáme dostatek dat pro příklon k jedné ze zkoumaných hypotéz**.

S ohledem na rozbor doporučení lze formulovat pokus o český překlad Wagenmakersova schématu v Tabulce 2.

Tabulka 2: Návrh slovního hodnocení hodnot Bayesova faktoru v češtině

Hodnota BF	Inverze BF	Navrhované české označení pro sílu evidence
Okolo 1	1	Nerohodná situace, nedostatek dat
1–3	1/3–1	Slabá, anekdotická
3–10	1/10–1/3	Středně silná
10–30	1/30–1/10	Silná
30–100	1/100–1/30	Velmi silná
>100	<1/100	Zcela přesvědčivá

Bayesův faktor pro dvouvýběrový t-test

Po teoretičtějších rozboru Bayesova faktoru a jeho možné interpretace se budeme věnovat jeho praktickému užití v analýze kvantitativních dat. Bude demonstrován výpočet BF pro nejčastěji používané procedury, tj. pro dvouvýběrový t-test, korelační koeficient a regresní analýzu (lineární model). Další situace jsou v zásadě analogické, a není tak třeba s nimi detailně čtenáře seznamovat s tím, že pro některé složitější procedury dosud možnost užívat bayesův faktor vůbec neexistuje, nebo není dosud implementován v softwaru. Kromě možnosti počítat bayesův faktor přímo z našich výzkumných dat existuje i možnost výpočtu bayesova faktoru z hodnot testovacích statistik (typicky tedy z již publikovaných výsledků), případně z hodnot informačního kritéria BIC. Pro všechny výpočty budeme užívat česká data z velkých mezinárodních sociologických šetření (ISSP a EVS), dále uvedené postupy je samozřejmě možné aplikovat i na data jiná (nicméně mělo by jít o data, která jsou reprezentativním výběrem, ideálně získaným procedurou náhodného výběru).

Aby bylo možné srovnat výsledek klasického postupu (dvouvýběrového t-testu s P hodnotou) a bayesova faktoru, provedeme výpočet nejdříve klasicky. Dvouvýběrový t-test použijeme pro srovnání délky vzdělání mužů a žen v ČR, výpočty budou založeny na datech

z poslední vlny šetření EVS 2017.²² Před samotným výpočtem zjistíme, zda jsou rozptyly délky vzdělání pro muže a ženy shodné (Leveneho test). Protože tomu tak není (výsledek zde neuvádíme), necháme vypočítat výsledky dvouvýběrového t-testu také dle Welcha (někdy jej označujeme jako robustní verzi dvouvýběrového t-testu). Výsledek zachycuje Tabulka 3.²³

Tabulka 3: Výsledek klasického t-testu pro srovnání délky vzdělání mezi muži a ženami v ČR (EVS 2017)

	t	df	p	Cohen's d
Student	2.760	1334	0.006	0.154
Welch	2.719	1069	0.007	0.153

$$n_1=528, n_2=808$$

Z tabulky snadno vyčteme (čteme druhý řádek, který odpovídá Welchovu testu), že hodnota testovací statistiky je $t=2,719$, což při 1069 stupních volnosti (df) vede k P hodnotě o velikosti 0,007. Prismaticem klasického doporučení, kdy výsledky s P hodnotami pod 0,05 považujeme za statisticky významné, bychom i zde sdělili, že rozdíl mezi délkou vzdělání mužů a žen je statisticky významný. V posledním sloupci Tabulky 3 je ovšem hodnota Cohenova d ,²⁴ která náš optimismus mírní. Hodnota činí 0,153, což při respektování Cohenových doporučení (Cohen 1988) znamená, že rozdíl je velmi malý (rozdíl v průměrné délce vzdělání mužů a žen činí 0,153 směrodatné odchylky této proměnné, což je opravdu poměrně málo). Výsledek je tedy dle P hodnoty zobecnitelný na populaci, ovšem věcně (pomocí Cohenova d) je efekt zanedbatelný. Nyní zkusíme pro stejnou úlohu vypočítat Bayesův faktor. Z úvodního ilustračního příkladu již víme, že pokud chceme užívat Bayesův faktor (i obecný bayesovský výpočet), musíme před samotným výpočtem specifikovat naši představu o výsledku, tzv. apriorní rozdělení. V případě dvouvýběrového t-testu se tak nabízí možnost specifikovat velikost Cohenova d . Software nabízí různé možnosti těchto specifikací (typicky se nabízí rovnoměrné rozdělení, normální rozdělení, polovina normálního rozdělení), nicméně pro t-test bývá nejběžnější užívat tzv. Cauchyho rozdělení. Jde v zásadě o rozdělení blízké normálnímu s tím, že jeho konce jsou těžší (tj. rozdělení méně rychle vpravo i vlevo klesá).²⁵ Toto

²² Šetření EVS 2017 probíhalo v ČR pod garancí prof. Rabušice z FSS MU. Český soubor obsahuje 1 812 respondentů starších 18 let.

²³ Výpočet byl proveden v softwaru JASP. Analogické výsledky by bylo samozřejmě možné získat ve velmi oblíbeném SPSS, ale i v jakémkoli jiném softwaru určeném pro statistické výpočty; bylo by možné použít i Excel. Všechny výpočty v tomto článku včetně datového souboru jsou online přílohou článku pro možnost jejich ověření či modifikace. JASP byl zvolen s ohledem na skutečnost, že obsahuje jak klasické, tak i bayesovské výpočty, jeho ovládání je velice snadné (analogické SPSS), umí mj. načítat data ve formátu pro SPSS a jde o produkt, který je dostupný zcela zdarma.

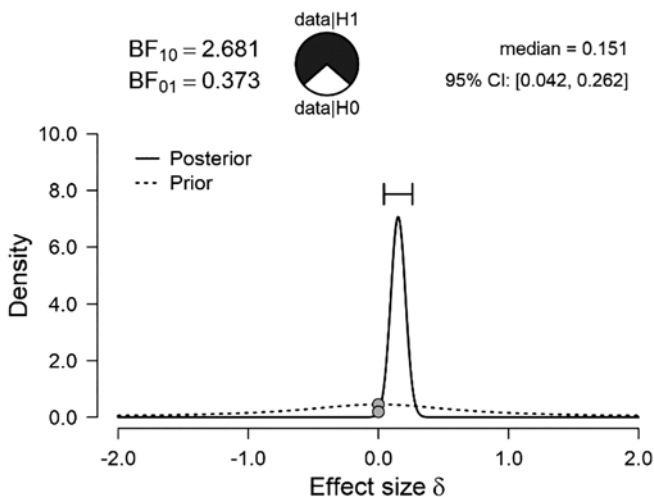
²⁴ Jde o nejužívanější míru věcné významnosti (*effect size*), více se o těchto mírách lze dozvědět v článku Soukupa (2013). Pro úplnost uvádíme, že Cohenovo d je vypočteno jako rozdíl průměrů v obou skupinách podělený směrodatnou odchylkou proměnné.

²⁵ Formálně jde o t rozdělení s jedním stupněm volnosti, tedy t rozdělení s nejtěžšími konci, jaké může mít.

rozdělení navrhl užívat sám Jeffreys a jde typicky o základní volbu v softwaru pro bayesovskou statistiku (JASP nevyjímaje). Cauchyho rozdělení má dva parametry, střed (*location*) a škálový parametr (*scale*). Nejčastěji, pokud nemáme žádné specifické očekávání, volíme střed v nule (tj. nulový rozdíl mezi skupinami považujeme za nejpravděpodobnější) a škálový parametr 0,707,²⁶ který odpovídá skutečnosti, že očekáváme velikost Cohena d v intervalu <-0,5;0,5>. Vyjádřeno slovníkem Cohena (1988) očekáváme, že rozdíl mezi skupinami bude maximálně střední.

Když využijeme toto apriorní rozdělení pro Cohenovo d, v našem případě získáme výsledky v Obrázku 3.

Obrázek 3: Bayesův faktor pro dvouvýběrový t-test (oboustranná hypotéza, JZS apriorní rozdělení)



Začneme nejdříve s hodnotou Bayesova faktoru. V Obrázku nalezneme BF₁₀ i BF₀₁. Je na nás, kterou hodnotu budeme vyhodnocovat, vybereme si BF₁₀, která porovnává alternativní hypotézu (rozdíl do cca 0,5 měřený Cohenovým d) s nulovou (nulový rozdíl). Hodnota BF₁₀ je 2,681, a pokud nahlédneme na dříve uvedená doporučení (Tabulky 1 a 2), není evidence ve prospěch alternativní hypotézy (ve srovnání s nulovou) nijak přesvědčivá. Vidíme zde jasný rozpor mezi výsledkem získaným klasicky (nízká P hodnota, statisticky významný výsledek) a bayesovsky (hodnota BF nedosahuje ani doporučené hodnoty 3). Ostatně na problematičnost klasicky získaného výsledku jsme již upozornili při interpretaci hodnoty Cohena d pro naše data (hodnota o velikosti 0,153 v Tabulce 3). Z tohoto pohledu se opět jeví jako smysluplné doporučení užívat kromě P hodnot též měř věcné významnosti (zde Cohena

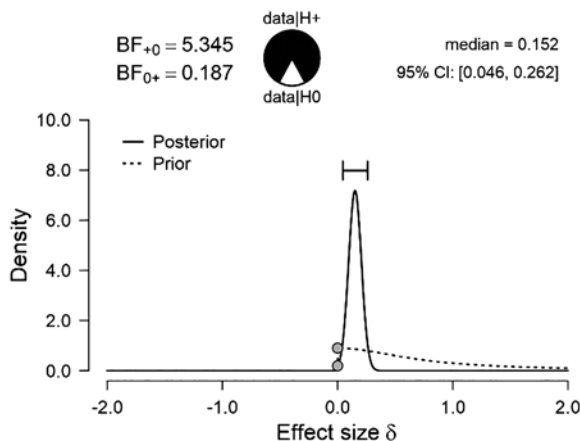
²⁶ Hodnota je odvozena od převrácené hodnoty druhé odmocniny ze 2. Posléze budeme toto rozdělení označovat jako JZS (srov. dále). Jeffreys původně navrhol rozdělení s hodnotou parametru 1, tedy ještě širší.

d). Vrátime se ale zpět k Bayesovu faktoru. Obrázek 3 nám nabízí i další výsledky, které mají vazbu k Bayesovu faktoru. Předně je nahoře uveden koláč, který graficky ukazuje poměr mezi pravděpodobnostmi alternativní (tmavá část) a nulové hypotézy. I když je tmavá výseč větší, není to nijak extrémní (je jen 2,7krát větší). Jinou možnost, jak získat hodnotu Bayesova faktoru, nabízí i graf zcela dole na Obrázku 3. Nejdříve si vysvětlíme logiku obou uvedených křivek (v obou případech jde o hustoty pravděpodobnosti, s ohledem na úspornost popisu to dále nebudeme uvádět). Tečkovaná plochá křivka je apriorním rozdělením Cohenova d (připomeňme, že vychází z Cauchyho rozdělení se středem v nule a škálovacím parametrem 0,707). Tím, že jsme získali data, jsme toto rozdělení „aktualizovali“ a získali jsme aposteriorní rozdělení (plná špičatá křivka). Výpočet Bayesova faktoru lze při bodové nulové hypotéze²⁷ získat pomocí Savage-Dickeyho hustoty tak, že v bodě 0 (na ose X) porovnáme výšku apriorního a aposteriorního rozdělení. V našem obrázku nelze toto určit přesně, ale platí, že apriorní rozdělení (čárkované) leží výše (přesně 2,681krát) než aposteriorní (plná křivka), a proto je $BF_{10}=2,681$.

Abychom demonstrovali, že hodnota Bayesova faktoru závisí na specifikaci apriorního rozdělení, uvedeme další dvě možnosti. Předpokládejme v nich:

- A. Muži mají na základě dřívější empirické evidence **vyšší** průměr délky vzdělání.
- B. Muži mají na základě dřívější empirické evidence **nižší** průměr délky vzdělání.

Obrázek 4: Bayesův faktor pro dvouvýběrový t-test (jednostranná hypotéza stranicí mužům, JZS apriorní rozdělení)



Naše apriorní rozdělení tedy v těchto případech budou odpovídat jen polovině Cauchyho rozdělení, které bude ale v bodě nula a jejím okolí vyšší. Výsledek našich analýz pro situaci A a B zachycují Obrázky 4 a 5. Obrázek 4 zachycuje výsledek pro alternativní hypotézu,

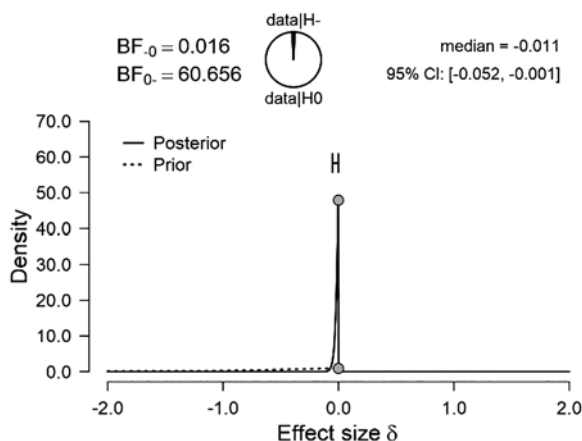
²⁷ Bodová nulová hypotéza tvrdí, že hodnota neznámého parametru (zde rozdílu mezi průměry) je rovna jediné konkrétní hodnotě (zde nule, tj. rozdíl mezi průměry neexistuje).

kteřá předpokládá, že muži mají v průměru delší vzdělání.²⁸ Hodnota Bayesova faktoru je nyní vyšší 5,345 a s ohledem na výše uvedená doporučení bychom tuto alternativní hypotézu favorizovali. Ostatně lze nahlédnout, že z koláče v Obrázku 4 skutečně ukrajuje alternativní hypotéza výrazně větší díl, stejně tak je patrné, že výška křivky apriorního rozdělení je výrazněji nad křivkou rozdělení aposteriorního. Hypotéza o vyšší průměrné délce vzdělání u mužů je tedy našimi daty podpořena.

Zcela jiný závěr zřejmě získáme, pokud budeme prověřovat jednostrannou alternativní hypotézu, která tvrdí, že ženy mají v průměru delší vzdělání. Bayesův faktor v tomto případě má hodnotu 0,016, jeho inverze pak 60,656. Jinými slovy nulová hypotéza (ženy mají v průměru stejné či kratší vzdělání) je oproti alternativní hypotéze našimi daty výrazně favorizována (připomeňme, že slovní vyjádření z Tabulky 2 by tuto evidenci označilo jako velmi silnou). Na Obrázku 5 je zcela zřejmé, že v bodě 0 je křivka apriorního rozdělení mnohokrát níže než křivka aposteriorního rozdělení.

Samozřejmě že v běžných situacích budeme mít typicky jedinou alternativní hypotézu. Pro zajištění maximální míry objektivity většinou vybíráme rozdělení specifikovaná statistiky, jehož příkladem bylo Cauchyho rozdělení (0;0,707).

Obrázek 5: Bayesův faktor pro dvouvýběrový t-test (jednostranná hypotéza stranící ženám, JZS apriorní rozdělení)

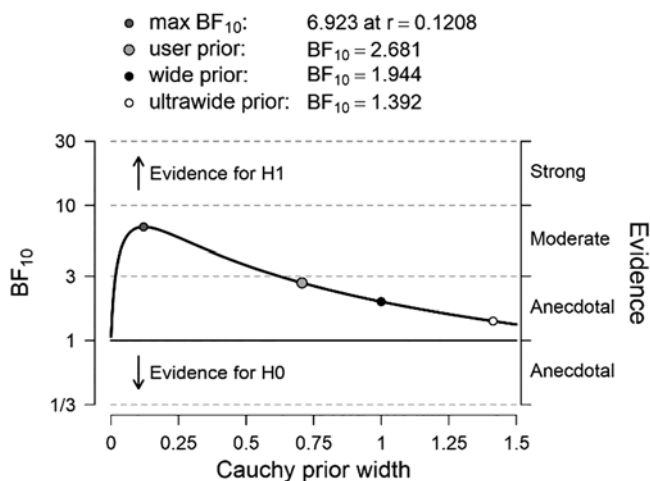


Často ale budeme specifikovat apriorní rozdělení precizněji, většinou na základě předchozích výsledků. Protože je tento postup ale do určité míry subjektivní, doporučuje se provádět tzv. analýzu citlivosti či robustnosti (*sensitivity analysis/robustness check*). Technicky to znamená, že zkusíme provést výpočet pro různá apriorní rozdělení a zjistit, jak se výsledky liší. Pokud minimálně, pak jsou naše výsledky nejspíš věrohodné, naopak při výrazných odlišnostech bychom měli být k našim výsledkům kritičtí. Doplňme, že tato následná analýza citlivosti

²⁸ Je vhodné upozornit, že v JASP je jednostranná alternativní hypotéza značena odlišně, buď znaménkem + (první skupina má vyšší průměr), nebo – (první skupina má nižší průměr).

je v oblasti baysovské statistiky doporučeným postupem. Depaoli a van de Schoot (2017) a mnozí další autoři se tímto doporučením řídí. Využijeme nabídky softwaru JASP, který umí základní analýzu citlivosti provést. Pokračujeme v našem příkladu s původním zadáním (tj. oboustrannou alternativní hypotézou specifikovanou pomocí Cauchyho rozdělení se škálovým parametrem o velikosti 0,707). Výsledek analýzy citlivosti zachycuje Obrázek 6.

Obrázek 6: Analýza citlivosti pro Bayesův faktor pro dvouvýběrový t-test (oboustranná hypotéza)



V obrázku je zachycena na ose X šířka apriorního rozdělení (naše původní odpovídala velikosti 0,707) a na svislé ose je pak zachycena hodnota BF. Snadno nahlédneme, že pro $r=0,707$ je označený šedý bod, který odpovídá námi spočtené hodnotě BF o velikosti 2,681. V obrázku jsou uvedeny pak ještě další hodnoty pro $r=1$ (označeno jako široké apriorní rozdělení), pro hodnotu 1,44 (označeno jako extrémně široké apriorní rozdělení) a dále je pak určena hodnota parametru Cauchyho rozdělení, pro kterou by hodnota BF byla největší (v našem případě je to 0,128). Pokud bychom tedy specifikovali naši alternativní hypotézu tak, že rozdíl mezi muži a ženami v délce vzdělání je velmi malý (cca do hodnoty 0,1 Cohenova d), získali bychom Bayesův faktor o velikosti 6,9, a takovou alternativní hypotézu bychom upřednostnili před nulovou.²⁹ Od této hodnoty směrem vpravo (tedy při širší specifikaci apriorního rozdělení³⁰) má hodnota Bayesova faktoru tendenci klesat. Nikdy se ovšem nedostane pod hodnotu 1, tj. při žádné (oboustranné) specifikaci apriorního rozdělení není možné získat evidenci ve prospěch nulové hypotézy. Za pozornost na Obrázku 6 stojí ještě popis svislé osy vpravo, kde je uplatněno Wagenmakersem upravené Jeffreysovo

²⁹ Musíme si ovšem uvědomit, že taková alternativní hypotéza není věcně nijak výrazně odlišná od nulové a nevede k nijak odlišnému konstatování, tj. pokud rozdíl existuje, je v zásadě zanedbatelný.

³⁰ Tedy očekávání, že rozdíl mezi skupinami je větší než 0,1 násobek směrodatné odchylky (tj. Cohenova d).

schéma (srov. Tabulku 1). Je z něj patrné, že při žádném apriorním rozdělení nepřekročíme úroveň středně silné evidence a pro běžné (opatrné, objektivní) apriorní rozdělení půjde jen o evidenci slabou (anekdotickou).

Bayesův faktor pro korelaci

Po rozboru dvouvýběrového t-testu (analogicky by se dala řešit situace pro jednovýběrový či párový t-test) se krátce zaměříme na korelaci. Pro jednoduchost vybereme Pearsonův korelační koeficient, postupy pro jiné koeficienty jsou analogické.

Opět začneme klasickým výpočtem s P hodnotou a poté výsledky srovnáme s Bayesovým faktorem. Pro ilustraci opět využijeme data z EVS 2017 za ČR a vypočteme korelaci pro vztah mezi délkou vzdělání a životní spokojeností. Výsledek klasického výpočtu uvádí Tabulka 4.

Tabulka 4: Výsledek klasického testu pro hodnocení korelace mezi délkou vzdělání a životní spokojeností v ČR (EVS 2017)

	Pearson's r	p
satis	0.088	0.001

n=1328

Hodnota korelace je věcně velmi malá (0,088), ale P hodnota 0,001 ukazuje na statistickou průkaznost našeho výsledku. Při respektování Cohenových doporučení (Cohen 1988) bychom konstatovali, že vztah mezi oběma veličinami je slabý.

Nyní vyhodnotíme korelaci pomocí Bayesova faktoru. Nejdříve musíme opět stanovit apriorní rozdělení korelačního koeficientu odpovídající alternativní hypotéze (nulová bude tvrdit, že veličiny spolu nesouvisí³¹). Pro první vzhled znovu zvolíme oboustrannou alternativní hypotézu. Jako základní nastavení se zpravidla (v softwaru) pro korelace používá posunutého beta rozdělení se šířkou 1.³² Toto rozdělení říká, že všechny hodnoty korelací mezi -1 a +1 jsou stejně pravděpodobné, a odpovídá tak rovnoměrnému rozdělení omezenému na tuto oblast. Parametr beta rozdělení můžeme měnit a tím ovlivnit naši představu, zatím však tuto možnost nevyužijeme. Výpočtem v JASP získáme Obrázek 7.

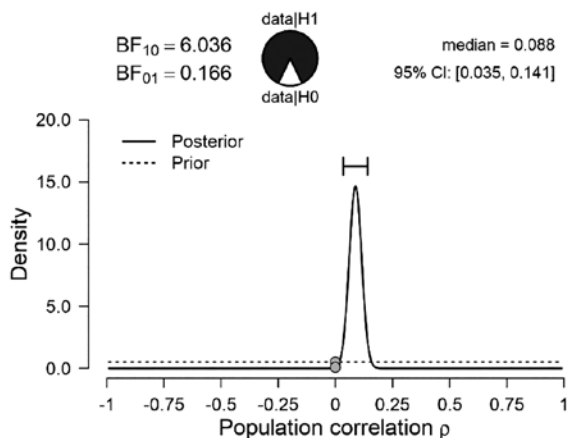
Z obrázku snadno zjistíme, že Bayesův faktor pro námi vybrané apriorní rozdělení korelace má hodnotu 6,³³ jde tedy o poměrně přesvědčivou evidenci ve prospěch alternativní hypotézy (tato je ve světle našich dat 6krát pravděpodobnější než nulová hypotéza hovořící o nezávislosti proměnných). Závěr je tedy analogický s testováním pomocí P hodnoty. Je namístě se nyní zamyslet nad získaným výsledkem. Naše korelace je věcně velmi slabá, přesto je však nespíše odlišná od nuly (jak s využitím P hodnoty, tak i pomocí BF).

³¹ Zcela shodně jako v klasickém testování pomocí P hodnoty.

³² Původní beta rozdělení je definováno pro hodnoty od 0 do +1, posunem vlevo o půl jednotky a jeho zdvojnásobením pak získáme rozdělení, které lze použít jako apriorní pro korelace.

³³ Připomeňme, že hodnotu bychom mohli odečíst též graficky, jde o poměr zobrazených hustot v bodě 0 (viz šedé body na Obrázku 7).

Obrázek 7: Bayesův faktor pro korelaci (oboustranná hypotéza, JZS apriorní rozdělení)

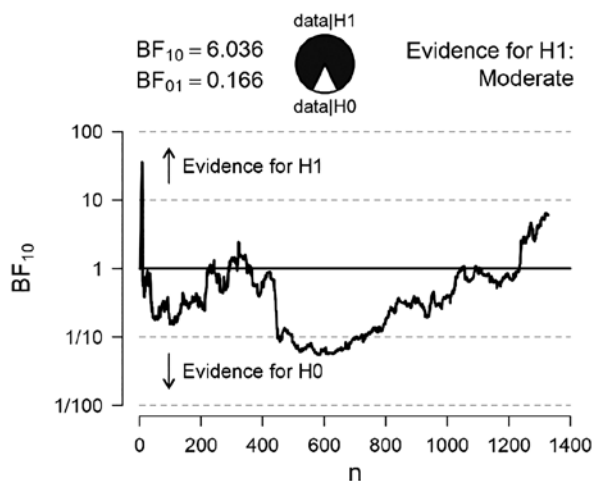


Důvody, proč preferujeme alternativní hypotézu před nulovou pomocí BF, jsou v našem případě zejména tyto:

1. Naše apriorní rozdělení korelace je zcela neinformativní.
2. Naše data jsou poměrně velká (soubor má více než 1800 respondentů).

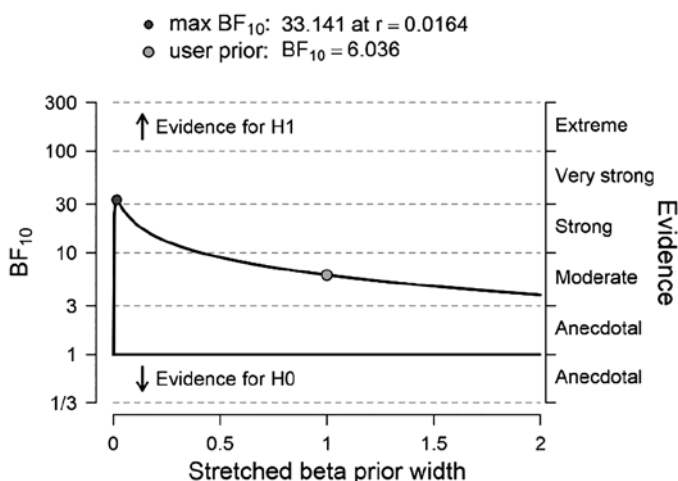
Je tedy patrné, že i Bayesův faktor je citlivý na velikost souboru (obdobně jako P hodnota) a se zvětšující se velikostí datového souboru jeho hodnota ceteris paribus se mění. Ukažme si tuto vlastnost na grafickém výstupu (Obrázek 8), který je označen jako sekvenční analýza (*sequential analysis*).

Obrázek 8: Sekvenční analýza pro Bayesův faktor pro korelaci (oboustranná hypotéza, JZS apriorní rozdělení)



Tento obrázek ukazuje, že pokud by náš soubor byl menší, pak by bylo namísto uvažovat o podpoře nulové hypotézy či o tom, že mezi hypotézami nelze rozhodnout (do cca 1200 jednotek). Poté však již začíná hodnota BF růst a začíná převažovat evidence ve prospěch alternativní hypotézy. Samozřejmě toto vše platí pro námi stanovené apriorní rozdělení (beta se šířkou 1), pro jiná rozdělení by byl obrázek jiný. Tento typ výstupu je vhodný pro úvahy, jak velké soubory potřebujeme pro již stabilní výsledky. Jde tedy o jistou analogii stanovení velikosti výběrového souboru pomocí síly testu či jiných používaných postupů. Mohli bychom samozřejmě též zkoumat vliv nastavení našeho apriorního rozdělení (tj. dopad různého parametru pro šířku beta rozdělení). Jde o již popsanou analýzu citlivosti. Její výsledek pro naši situaci zachycuje Obrázek 9.

Obrázek 9: Analýza citlivosti pro Bayesův faktor pro korelaci (oboustranná hypotéza)



Je patrné, že nejvyšší hodnotu Bayesova faktoru bychom získali pro velice malou šířku apriorního rozdělení (tj. dle alternativní hypotézy bychom očekávali, že hodnota korelace je velice blízká nule³⁴), naopak nejnižší hodnota BF by byla pro velice široce rozevřené apriorní rozdělení (na obr. 9 je maximem hodnota 2), které předpokládá, že nejpravděpodobnější jsou velké korelace a jiné hodnoty jsou málo pravděpodobné.

Samozřejmě opět platí, že volba apriorního rozdělení by měla respektovat naši znalost problému. Pokud v zásadě nic nevíme, je vhodné užít nastavení, které předpokládá stejnou pravděpodobnost úrovně korelace napříč jejími možnými hodnotami. Nicméně většinou minimálně tušíme, zda je očekávaná korelace pozitivní či negativní, pak bychom použili jen polovinu rozdělení. A samozřejmě navíc můžeme specifikovat, zda jsou pravděpodobnější větší hodnoty (parametr šířky větší než 1), nebo naopak menší hodnoty korelace (parametr menší než 1).

³⁴ Alternativní hypotéza by tedy byla velice podobná nulové. O této situaci jsme již hovořili u t-testu.

Bayesův faktor pro regresní analýzu

Poslední procedurou, na které budeme demonstrovat výpočetní aspekty Bayesova faktoru, bude regresní analýza, přesněji řečeno lineární regresní analýza. Připomeňme, že zejména v sociologii jde o jednu z nejužívanějších analytických technik (Bernardi et al. 2017). Využití Bayesova faktoru zde bude komplikovanější než pro přechodí situace. Zatímco v t-testu či korelaci máme dvě jasně proti sobě stojící hypotézy, které se zaměřují na (ne) existenci (a velikost) očekávaného rozdílu, resp. souvislosti, u regresní analýzy může být konkurujících hypotéz mnohem více. Spíše než o konkurenčních hypotézách hovoříme o konkurenčních modelech (ostatně bayesiánci tento výraz upřednostňují). Počet těchto modelů, které můžeme srovnávat, je odvozen od toho, kolik máme nezávisle proměnných v regresním modelu. Pokud užíváme jednoduchou regresní analýzu (jedna nezávisle proměnná), pak je situace analogická ke korelaci a budeme srovnávat model, který předpokládá nezávislost (obsahuje jen konstantu, tzv. nulový model), a model, který předpokládá závislost (kromě konstanty obsahuje též nezávisle proměnnou). V sociálních vědách je ale preferována regrese mnohonásobná s více prediktory. A zde kromě modelu nezávislosti můžeme mít model s prvním prediktorem, druhým, třetím atd. a jakoukoli jejich kombinací. Obecně platí, že těchto modelů, které můžeme srovnávat pomocí Bayesova faktoru, je určen počtem prediktorů (k) jako 2^k . Nadto je možné tento počet ještě zvýšit, pokud bychom uvážili, že některé prediktory můžeme kombinovat do interakcí, některé prediktory působí nelineárně (např. ve druhé či třetí mocnině apod.). Prakticky se tento problém řeší většinou tak, že při výpočtech Bayesova faktoru počítáme Bayesův faktor, který srovnává každý jednotlivý model s modelem obsahujícím pouhou konstantu (nulový model), nebo srovnáváme každý model s modelem, který je nejlepší (typicky se určuje jako model s nejvyšší aposteriori pravděpodobností). Při větším počtu modelů (typicky od řádu desítek) pak většinou stanovujeme hodnotu BF jen pro několik málo (cca 10) nejlepších modelů, doplnkově někdy sledujeme též modely nejhorší. Musíme si uvědomit, že P hodnoty jsou v regresní analýze k uživatelům přátelštější, je jich vždy jen tolik, kolik je námi odhadovaných parametrů.

Kromě obtížnosti s počtem konkurenčních modelů musíme pro odhad BF ještě řešit apriorní nastavení. Opět zde situace není snadná, protože pro bayesiánský odhad regresního modelu je třeba stanovit apriorní rozdělení každého regresního koeficientu (včetně konstanty), a dokonce též rozptylu náhodné složky. Zde si většinou pomáháme přednastavenými možnostmi, zřejmě nejčastější je využití tzv. JZS apriorního rozdělení.³⁵ Pro toto rozdělení je třeba nastavit parametr r , jeho hodnota bývá typicky přednastavena na velikost 0,354 (např. JASP či balíček BayesFactor v R, ze kterého JASP čerpá).

³⁵ Jde o Jeffreys–Zellner–Siowovo apriorní rozdělení (Jeffreys, 1961; Zellner & Siow, 1980). Zellner se Siowem modifikovali původní Jeffreysův návrh a v současnosti jde o typické nastavení v rámci výpočtu BF v regresních modelech. Dodejme, že ale zdaleka ne jediné, např. software JASP disponuje v regresní analýze devíti možnými apriorními rozděleními. Připomeňme, že s JZS rozdělením jsme již pracovali v rámci t-testu i korelace, aniž bychom to zdůrazňovali.

Pro ilustraci zvolíme regresní model pro příjem,³⁶ jako nezávisle proměnné vybereme délku vzdělání, pohlaví, věk a počet hodin práce za týden. Využijeme jeden z nejnovějších datových souborů ISSP 2019.³⁷ Po nezbytné přípravě proměnných³⁸ je možné vypočítat regresní model klasicky a poté se zaměřit na bayesovský výpočet a zejména pak na Bayesův faktor.

Klasickým výpočtem získáme výsledek v Tabulce 5.

Tabulka 5: Výsledek klasické regresní analýzy: závisle proměnná čistý osobní měsíční příjem v ČR (ISSP 2019)

Model		b	S.E.	Beta koef.	t	p
1	(Intercept)	-1045.46	3616.09		-0.289	0.773
	muz	4163.88	986.86	0.172	4.219	< .001
	vek	66.68	41.46	0.065	1.608	0.108
	vzdel	1315.61	183.86	0.289	7.156	< .001
	hodiny	137.89	40.99	0.136	3.364	< .001

$R_2=0,138$, $F(4,532)=21,33$ ($P<0,001$), $n=537$

Náhledem na výsledky můžeme zjistit, že P hodnoty u jednotlivých prediktorů jsou poměrně malé ($<0,001$), výjimkou je proměnná věk, kde P hodnota 0,108 převyšuje konvenčně užívanou 5% hladinu statistické významnosti a vliv věku na příjem bychom tedy pomocí klasického testování (díličními t-testy) neprokázali. Doplňme, že i věcně je vliv věku poměrně malý (srov. zejména hodnotu beta koeficientu, která je pro věk zdaleka nejnižší).

Pro bayesovské výpočty bude potřebné zvolit apriorní rozdělení (využijeme zmíněné JZS). Dále již dopředu víme, že bayesovský výpočet bude srovnávat jednotlivé modely obsahující různé kombinace nezávisle proměnných. Při čtyřech prediktorech tak získáme 16 modelů (2^4) a pro všechny tyto modely se vypočte BF srovnávající konkrétní model s nulovým modelem či nejlepším modelem (jde o model s nevyšší aposteriorní pravděpodobností). Plný bayesovský výpočet také poskytne odhad jednotlivých koeficientů modelu se všemi prediktory (typicky pomocí průměru jejich aposteriorních rozdělení), jejich kredibilitní

³⁶ Pro korektní analýzu by bylo vhodné původní proměnnou logaritmovat, protože rozdělení příjmů je typicky sešikmené. Praktické analýzy (jak v sociologii, tak ekonometrii) ale ukazují, že lineární model pro příjmy je přijatelným zjednodušením, a regresní diagnostika (typicky založená na analýze reziduí) tuto skutečnost běžně potvrzuje (není zde uvedena).

³⁷ Autor tímto děkuje Daně Hamplové ze Sociologického ústavu AV ČR za možnost tato data využít dříve, než budou uvolněna odborné veřejnosti k dalším analýzám (uvolnění je plánováno na září 2020).

³⁸ Byla vybrána jen skupina osob pracujících. Dále byly kromě odstranění chybějících hodnot též eliminovány extrémní příjmy (výrazně nízké a vysoké) a extrémní počty hodin práce (výrazně nízké a vysoké) a proměnná zachycující pohlaví byla rekódována na hodnoty 0 (žena) a 1 (muž). Tato úprava byla provedena zejména s ohledem na další možné využití interakčních členů v modelu (srov. dále).

intervaly (bayesiánskou obdobu intervalů spolehlivosti) a případně i další charakteristiky. S ohledem na skutečnost, že článek se zaměřuje jen na Bayesův faktor, zde tyto výstupy „plnohodnotné“ bayesovské statistiky neuvádíme a nebudeme jim věnovat pozornost. Pokud necháme zobrazit výstupy pro všechny potenciální modely a provedeme srovnání BF vůči nejlepšímu modelu, získáme Tabulku 6.³⁹

Tabulka 6: Srovnání regresních modelů pomocí Bayesova faktoru: závisle proměnná čistý osobní měsíční příjem v ČR (ISSP 2019)

Model	P(M)	P(M data)	BF ₁₀	R ²
muz + vek + vzdel + hodiny	0.200	0.672	1.000	0.138
muz + vzdel + hodiny	0.050	0.315	1.876	0.134
muz + vzdel	0.033	0.007	0.066	0.116
muz + vek + vzdel	0.050	0.005	0.028	0.120
vek + vzdel + hodiny	0.050	2.161e-4	0.001	0.109
vzdel + hodiny	0.033	1.432e-4	0.001	0.103
vzdel	0.050	2.085e-6	1.240e-5	0.079
vek + vzdel	0.033	1.061e-6	9.467e-6	0.086
muz + hodiny	0.033	1.241e-10	1.107e-9	0.053
muz + vek + hodiny	0.050	5.258e-11	3.129e-10	0.055
muz	0.050	4.430e-12	2.636e-11	0.032
hodiny	0.050	1.083e-12	6.446e-12	0.027
muz + vek	0.033	6.458e-13	5.764e-12	0.034
vek + hodiny	0.033	2.817e-13	2.514e-12	0.031
Null model	0.200	3.859e-14	5.741e-14	0.000
vek	0.050	2.453e-15	1.460e-14	0.004

Stručně popišme logiku jednotlivých sloupců Tabulky 6. Jednotlivé řádky jsou jednotlivé modely, s plusem jsou uvedeny prediktory využití v těchto modelech. V prvním sloupci (P(M)) jsou apriorní pravděpodobnosti jednotlivých modelů. JASP⁴⁰ je stanovuje tak, že nejdříve rozdělí jednotkovou pravděpodobnost podle počtu prediktorů. V našem případě nemáme buď žádný (jen konstantu), nebo jeden až čtyři, tedy celkem 5 možností (pravděpodobnost pro jednu úroveň prediktorů je tedy 0,2). Pro každý počet prediktorů (0–4) pak ještě rozdělí pravděpodobnost na jednotlivé modely se stejným počtem prediktorů. Proto má model konstanty (*null model*) apriorní pravděpodobnost 0,2, obdobně jako model se všemi prediktory na prvním řádku. Naopak například modely pro jediný prediktor mají každý apriorní pravděpodobnost jen 0,05 (apriorní pravděpodobnost 0,2 se dělí mezi čtyři modely pro jednotlivé osamělé prediktory).

³⁹ Připomínáme, že tabulka se většinou zkracuje na několik málo nejlepších (nejhorších modelů). Zde pro ilustraci máme tabulku kompletní, v dalším příkladu už provedeme redukci.

⁴⁰ Resp. balíček BayesFactor v R, ze kterého JASP čerpá.

Další sloupec pak určuje aposteriorní pravděpodobnosti jednotlivých modelů, symbolicky značených jako $P(M/data)$. Dodejme, že dle aposteriorních pravděpodobností jsou modely v tabulce tříděny (od největší po nejmenší). Předposlední sloupec pak ukazuje hodnoty BF pro srovnání mezi nejlepším modelem a konkrétním modelem, poslední sloupec pak udává hodnoty R^2 (jde o běžně známé hodnoty z klasické regresní analýzy, zde se žádný bayesiánský výpočet neuplatní).

Zaměříme se na hodnoty BF. Pro model na prvním řádku (obsahuje všechny prediktory) je hodnota jednotková (srovnává model se sebou samým). Pro druhý model (neobsahuje proměnnou věk) je hodnota BF 1,9. Naznačuje nám, že model bez vlivu věku je ve světle našich dat mírně lepší než model, který s touto proměnnou pracuje. Jinými slovy tato proměnná asi nic zvláštního modelu nepřidá. Zde připomeňme základní princip pro modelování ve statistice, tzv. princip parsimonie. Máme-li dva či více modelů obdobné kvality, volíme model jednodušší. Náš výsledek nám tedy naznačuje, že proměnnou věk bychom mohli bez ztráty kvality z modelu vypustit. Ostatně můžeme srovnat hodnoty R^2 obou modelů (0,138 vs. 0,134) a můžeme se zpětně podívat i na Tabulku 5 a poměrně vysokou P hodnotu pro proměnnou věk. Pokud nahlédneme hodnoty BF v dalších řádcích, snadno zjistíme, že jsou velmi malé (nejvyšší je 0,066⁴¹ pro model obsahující vliv pohlaví a vzdělání) a že další redukce modelu již není žádoucí, protože model obsahující všechny proměnné je ve světle našich dat výrazně lepší. K těmto srovnáním ještě přidejme poznámku navazující na text z úvodu článku. Stejně jako můžeme stanovit BF pro srovnání dvou modelů, můžeme též počítat poměr dvou hodnot BF, pokud jsou vztaženy ke stejnému srovnávacímu modelu (u nás nejlepší). Pokud tedy budeme chtít spočítat BF pro srovnání modelu obsahujícího pohlaví se vzděláním s modelem, který má navíc ještě počet hodin práce (řádky 3 a 2 Tabulky 6), pak podělíme příslušné hodnoty BF. Získáme tak hodnotu 0,0352 (=0,066/1,876). Určitě tedy budeme preferovat model, který obsahuje i počet hodin v práci, jeho pravděpodobnost ve světle našich dat je cca 28krát vyšší.⁴²

Před opuštěním regresní ukázky si ještě položíme otázku, zda by nebylo smysluplné přidat některý interakční člen, tj. kromě jednotlivých prediktorů zohlednit též jejich kombinace. Nabízí se například úvaha o různém vlivu délky vzdělání na příjem pro muže a ženy, tedy přidání interakčního členu proměnné vzdělání a pohlaví. Opět nejdříve uvedme klasický výpočet a srovnáme ho s výpočtem bayesovským. Pro zjednodušení nebudeme již do výpočtů zahrnovat proměnnou zjišťující věk respondenta. Výsledky klasického výpočtu zachycuje Tabulka 7.

⁴¹ Jde tedy o hodnotu menší než 0,1 (srov. Tabulky 1 a 2), evidence tedy jasně preferuje komplexní model.

⁴² Hodnotu 28 jsme získali jako převrácenou hodnotu námi vypočtené hodnoty BF (0,0352).

Tabulka 7: Výsledek klasické regresní analýzy: závisle proměnná čistý osobní měsíční příjem v ČR, model s interakcí (ISSP 2019)

Model		b	S.E.	Beta koef.	t	p
1	(Intercept)	7968.6	3562.51		2.237	0.026
	muz	-6601.3	4952.87	-0.273	-1.333	0.183
	vzdel	942.7	245.10	0.207	3.846	< .001
	hodiny	110.2	36.02	0.123	3.058	0.002
	muz * vzdel	830.6	367.78	0.465	2.258	0.024

$R^2=0,141$, $F(4,539)=22,08$ ($P<0,001$), $n=544$

Přidaný interakční člen má P hodnotu o velikosti 0,024, prohlásili bychom jej tedy za statisticky průkazný. Všimněme si též skutečnosti, že beta koeficient pro tuto interakci je nejvyšší ze všech. Za připomínku stojí i skutečnost, že proměnná charakterizující pohlaví má nyní P hodnotu vysokou (jde o klasický efekt při přidávání interakcí), nicméně volíme zde převažující hierarchický postup,⁴³ tj. při přidávání interakcí ponecháváme i proměnné (hlavní efekty), z nichž interakce vznikly. Nyní se zaměříme na model s interakcí bayesovsky, pro zjednodušení si zobrazíme jen srovnání modelu s interakcí a bez ní (Tabulka 8).

Tabulka 8: Srovnání regresních modelů pomocí Bayesova faktoru pro model s interakcí: závisle proměnná čistý osobní měsíční příjem v ČR (ISSP 2019)

Model	P(M)	P(M data)	BF ₁₀	R ²
muz * vzdel	0.500	0.641	1.000	0.141
Null model (incl. muz, vzdel, hodiny)	0.500	0.359	0.560	0.133

V Tabulce 8 je pro nás relevantní jen jediná hodnota BF, ta ve druhém řádku. Činí 0,56, tj. ve světle našich dat je model bez interakce zhruba 2krát horší. My ale již víme, že tato hodnota není nijak přesvědčivá a nevede k jasné preferenci modelu s interakcí. Zatímco P hodnota se tedy vyjadřovala jasně pro zařazení interakce, pomocí hodnoty BF tuto podporu nenacházíme (je velice slabá). Opět tedy (jako v případě t-testu) dospíváme k rozporu mezi tím, co naznačuje P hodnota, a hodnota BF. Ostatně to, že P hodnoty mají příliš velkou tendenci k zamítání nulových hypotéz, je statistikům známo velmi dlouho, zřejmě nejslavnější text s matematickým důkazem této vlastnosti napsali Sellke, Bayarri a Berger (2001).

⁴³ Výraz hierarchický postup má v regresní analýze ještě další dva významy. První (často využívaný psychology) označuje skutečnost, že postupně zařazujeme různé skupiny proměnných a modely srovnáváme (typicky pomocí R^2), druhý pak označuje hierarchičnost v datech (například data jsou za žáky, kteří jsou hierarchicky zařazení do škol), kterou je nutno zohlednit jinými výpočetními postupy (označovanými obecně jako víceúrovňové modelování).

Závěrem aneb přednosti a nedostatky Bayesova faktoru

Z předchozího textu poměrně jasně vyplývá, že i když v některých momentech je Bayesův faktor lepší než P hodnota (typicky tolik nepřeceňuje naše výsledky, umožňuje vyjádřit podporu nulové hypotéze), ani tento nástroj není zcela bez problémů. Ostatně alternativ k P hodnotě existuje veliké množství. Jako jedna z prvních alternativ bylo doporučeno užívat intervaly spolehlivosti (Cumming a Finch 2005), dále míry věcné významnosti (Cohen 1988), bootstrap (Efron 1979) nebo informační kritéria typu BIC (Raftery 1995). V poslední době se objevují další možné alternativy. Za zmínku stojí P-křivky (Simonsohn et al. 2014), TOST (Lakens et al. 2018), SGPV⁴⁴ (Blume et al. 2018) a FPR⁴⁵ (Colquhoun 2019).

Obecně lze konstatovat, že hledání univerzálního a pro všechny situace vhodného statistického nástroje připomíná hledání pověstného svatého grálu. Lze předvídat, že nikdy nebude nalezen. Ostatně pro různé výzkumné situace se hodí jiné postupy a logicky i jiné nástroje je třeba využít pro vyhodnocení výsledků. Bayesův faktor je jeden z těch, který některé přednosti nabízí a pro některé situace je vhodnější než P hodnota, ale rozhodně nejde o jediný univerzální nástroj. Pro hodnocení jeho předností a nedostatků ve srovnání s P hodnotou můžeme využít 5 kritérií stanovených Wagenmakersem (2007: 795):

1. Statistický postup by měl záviset jen na datech, která jsme reálně napozorovali.
2. Výsledky statistického postupu by neměly nijak odrážet záměry výzkumníka.
3. Statistický postup by měl zohledňovat míru důkazů ve prospěch nulové i alternativní hypotézy.
4. Statistický postup by měl být snadno aplikovatelný (spočítán v ruce či pomocí dostupného programu).
5. Statistický postup by měl být pokud možno objektivní, tj. různí výzkumníci by pro stejná data a stejné modely měli získat stejné výsledky.

Prizmatem těchto kritérií je zřejmé, že P hodnota nespňuje první tři kritéria, naopak naplňuje poslední dvě kritéria. Bayesův faktor naopak bez problémů splňuje kritéria 1–3, s komplikacemi naplňuje i čtvrté,⁴⁶ problémem tak zůstává páté kritérium. Bayesův faktor je tedy vždy do určité míry subjektivní. Připomeňme, že míra subjektivity v bayesiánském uvažování je samým základem této koncepce. Vždyť pro výpočty musíme dopředu stanovit apriorní rozdělení a to ovlivní mj. i výpočty hodnot BF. Bayesiánci se ale dlouhodobě snaží s tímto problémem vypořádat. Daří se jim to jen částečně. Hlavními nástroji jsou jednak různá doporučená apriorní rozdělení, která jsou dostatečně široká (hovoří se díky tomu o tzv. objektivním bayesiánství⁴⁷). Příkladem je zejména opakovaně zmíněné JZS rozdě-

⁴⁴ Second Generation of P Values.

⁴⁵ False positive risk, charakteristika užívaná typicky v teorii rozhodování pro vyhodnocení úspěšnosti našeho klasifikátoru.

⁴⁶ Wagenmakers v době psaní článku (2006) konstatoval, že BF zatím toto kritérium nenaplňuje. Ale byl to sám Wagenmakers, kdo inicioval vytvoření softwaru JASP, napsal desítky článků a zasloužil se tak o obrovský posun v dané oblasti.

⁴⁷ Druhou větví je pak subjektivní bayesiánství, které tento postup striktně odmítá a považuje za přednost skutečnost, že každý výzkumník určí apriorní rozdělení dle svých znalostí subjektivních přesvědčení.

lení. Druhým postupem je pak analýza citlivosti, tj. určení dopadu apriorních nastavení na výsledky. Tento postup se stává v zásadě standardem práce v oblasti bayesiánské analýzy dat. Ostatně právě plnohodnotná bayesiánská analýza dat je zřejmě odpovídající alternativou k P hodnotě, která má zejména mezi předními statistiky mnoho zastánců. V tomto článku jsme tak pomyslně zůstali na půl cesty. Opustili jsme klasické statistické výpočty (statistické testování založené na P hodnotě) a využili některé postupy bayesovské statistiky (hodnotu Bayesova faktoru). Je korektní přiznat, že pravověrní bayesiánci však tento nástroj téměř ignorují a postupují mnohem složitěji. Ostatně tomuto fenoménu jsou dnes věnovány celé knihy, zmiňme některé z nich (Lee 2012; Lee a Wagenmakers 2013; Kruschke 2014; Kaplan 2014; Gelman et al. 2013). Na otázku, zda je Bayesův faktor lepší než P hodnota, lze zřejmě odpovědět podmíněným ano. V některých situacích tomu tak je. Je ale bohužel i mnoho situací, kdy nám ani P hodnota, ani Bayesův faktor nepomohou. Máme-li dát na konci článku praktické doporučení, pak je jistě vhodné společně s P hodnotami uvádět i hodnotu bayesova faktoru, nebo alespoň publikovat takové charakteristiky, z nichž lze Bayesův faktor určit. Další výzkumníci pak mohou výpočty provést za nás a případně je využít pro další postupy, jako je například metaanalýza.

Literatura

- AERA. 2006. „Standards for Reporting on Empirical Social Science Research in AERA Publications.“ *Educational Researcher* 35(6): 33–40.
- APA. 2010. *Publication Manual of the American Psychological Association*. 6th Edition. Washington, DC: American Psychological Association.
- BAYES, Thomas a R. PRICE. 1763. „An Essay towards Solving a Problem in the Doctrine of Chance. By the Late Rev. Mr. Bayes, Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S.“ *Philosophical Transactions of the Royal Society of London* 53: 370–418.
- BERNARDI, Fabrizio, Lela CHAKHAIA a Liliya LEOPOLD. 2017. „Sing Me a Song with Social Significance: The (Mis)Use of Statistical Significance Testing in European Sociological Research.“ *European Sociological Review* 33(1): 1–15.
- BLUME, Jeffrey D., Lucy D'AGOSTINO MCGOWAN, William D. DUPONT a Robert A. GREEVY Jr. 2018. „Second Generation p-Values: Improved Rigor, Reproducibility, & Transparency in Statistical Analyses.“ *PLoS ONE* 13(3): e0188299.
- COHEN, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition. Lawrence Erlbaum Associates.
- COLQUHOUN, David. 2019. „The False Positive Risk: A Proposal Concerning What to Do About *p*-Values.“ *The American Statistician* 73(sup1): 192–201.
- CUBEREK, Roman a Karel FRÖMEL. 2011. „K problematice výzkumného výběru a testování nulové hypotézy.“ *Československá psychologie* 55(5): 468–477.
- CUMMING Geoff a Sue FINCH. 2005. „Inference by Eye: Confidence Intervals and How to Read Pictures of Data.“ *American Psychologist* 60(2): 170–180.
- DEPAOLI, Sarah a Rens van de SCHOOT. 2017. „Improving Transparency and Replication in Bayesian Statistics: The WAMBS-Checklist.“ *Psychological Methods* 22: 240–261.
- EFRON, Bradley. 1979. „Bootstrap Methods: Another Look at the Jackknife.“ *Annals of Statistics* 7(1): 1–26.

- GELMAN, Andrew, John B. CARLIN, Hal S. STERN, David B. DUNSON, Aki VEHTARI a Donald B. RUBIN, 2013. *Bayesian Data Analysis*. Third Edition. Boca Raton, FL: CRC Press.
- GIGERENZER, Gerd. 2018. „Statistical Rituals: The Replication Delusion and How We Got There.“ *Advances in Methods and Practices in Psychological Science* 1(2): 198–218.
- GOOD, Irving J. 1979. „Studies in the History of Probability and Statistics. XXXVII A. M. Turing’s Statistical Work in World War II.“ *Biometrika* 66(2): 393–396.
- HALLER, Heiko a Stefan KRAUSS. 2002. „Misinterpretations of Significance: A Problem Students Share with Their Teachers?“ *MPR-online* 7(1): 1–20.
- JEFFREYS, Harold. 1939. *Theory of Probability*. 1st Edition. Oxford: The Clarendon Press.
- JEFFREYS, Harold. 1948. *Theory of Probability*. 2nd Edition. Oxford: The Clarendon Press.
- JEFFREYS, Harold. 1961. *Theory of Probability*. 3rd Edition. Oxford: Oxford University Press.
- KAPLAN, David. 2014. *Bayesian Statistics for the Social Sciences*. New York: Guilford Publications.
- KASS, Robert E. a Adrian E. RAFTERY. 1995. „Bayes Factors.“ *Journal of the American Statistical Association* 90(430): 773–795.
- KRUSCHKE, John. 2014. *Doing Bayesian Data Analysis. A Tutorial with R, JAGS, and Stan*. 2nd Edition. Cambridge, MA: Academic Press.
- LAKENS, Daniël, Anne M. SCHEEL, Peder M. ISAGER. 2018. „Equivalence Testing for Psychological Research: A Tutorial.“ *Advances in Methods and Practices in Psychological Science* 1(2): 259–269.
- LEE, Michael D. a Eric-Jan WAGENMAKERS. 2013. *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.
- LEE, Peter M. 2012. *Bayesian Statistics: An Introduction*. Hoboken, NJ: Wiley.
- OAKES, M. W. 1986. *Statistical Inference: A Commentary for the Social and Behavioural Sciences*. Hoboken, NJ: Wiley.
- RAFTERY, Adrian. 1995. „Bayesian Model Selection in Social Research.“ *Sociological Methodology* 25: 111–163.
- SELLKE, Thomas, M. J. BAYARRI a James O. BERGER. 2001. „Calibration of p Values for Testing Precise Null Hypotheses.“ *The American Statistician* 55: 62–71.
- SIMONSOHN, Uri, Leif D. NELSON a Joseph P. SIMMONS. 2014. „P-Curve: A Key to the File-Drawer.“ *Journal of Experimental Psychology: General* 143(2): 534–547.
- SOUKUP, Petr. 2010. „Nesprávná užívání statistické významnosti a jejich možná řešení.“ *Data a výzkum – SDA Info* 4(2): 77–104.
- SOUKUP, Petr. 2013. „Věcná významnost výsledků a její možnosti měření.“ *Data a výzkum – SDA Info* 7 (2): 125–148.
- SOUKUP, Petr. 2019. „P a d (Používání statistické a věcné významnosti v českých sociálních vědách).“ *Sociologický časopis* 55(2): 215–254.
- SOUKUP, Petr a Ladislav RABUŠIC. 2007. „Několik poznámek k jedné obsesi českých sociálních věd – statistické významnosti.“ *Sociologický časopis* 43(2): 379–395.
- WAGENMAKERS, Eric-Jan. 2007. „A Practical Solution to the Pervasive Problems of p Values.“ *Psychonomic Bulletin & Review* 14: 779–804.
- WASSERMAN, Larry. 2004. *All of Statistics (A Concise Course in Statistical Inference)*. New York: Springer.
- WETZELS, Ruud, Dora MATZKE, Michael D. LEE, Jeffrey N. ROUDER, Geoffrey J. IVERSON, Eric-Jan WAGENMAKERS. 2011. „Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 *t* Tests.“ *Perspectives on Psychological Science* 6: 291–298.
- ZELLNER Arnold a Aloysius SIOW. 1980. „Posterior Odds Ratios for Selected Regression Hypotheses.“ *Trabajos de Estadística Y de Investigación Operativa* 31: 585–603.

Autor

Petr Soukup působí na katedře sociologie Institutu sociologických studií FSV UK. Ve výuce se věnuje nejvíce statistice a analýze dat, v sociálněvědní komunitě aktivně propaguje skrze semináře a letní školy užívání novějších postupů zejména z oblasti vícerozměrné statistiky. Vědecky se zaměřuje zejména na oblast sociologie vzdělání. Je spoluautorem úspěšné učebnice sociálněvědní analýzy dat v SPSS, nyní dokončuje pro vydání pokračování této učebnice a verzi učebnice pro software R (obojí bude vydáno v nakladatelství Masarykovy univerzity). Je autorem pěti článků, systematicky pojednávajících o P hodnotě a jejím užívání v oblasti sociálních věd, v návaznosti na tyto texty v současnosti dokončuje monografii o tomto tématu (vydání je naplánováno v nakladatelství Karolinum).

Kontakt: petr.soukup@fsv.cuni.cz