



# **PŘÍRUČKA EXPERIMENTÁLNÍ EKONOMIE**

Vojtěch Zíka





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



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

# PŘÍRUČKA EXPERIMENTÁLNÍ EKONOMIE

Vojtěch Zíka

## **PŘÍRUČKA EXPERIMENTÁLNÍ EKONOMIE**

Copyright © Vojtěch Zíka, 2023

Editace © Sára Laníková, 2023

ISBN 978-80-7561-410-0

Vydaly Laboratoře behaviorálních studií Univerzity Jana Evangelisty Purkyně v Ústí nad Labem roku 2023. Všechna práva vyhrazena. Žádná část této publikace nesmí být kopírována a rozmnožována bez písemného souhlasu autora.

Tato publikace vznikla v rámci projektu OPVVV "Smart City-Smart Region-Smart Community" (CZ.02.1.01/0.0/0.0/17\_048/0007435) a projektu OPVVV "Centrum regulace a behaviorálních studií pro výzkumně zaměřené studijní programy UJEP"(CZ.02.1.01/0.0/0.0/16\_017/0002689).

# Obsah

Úvod	1
<b>1 Odměňování participantů</b>	<b>15</b>
1.1 Úloha incentív	15
1.2 Zákaz klamání participantů	18
1.3 Prvotní vybavení	19
1.4 Shrnutí	23
<b>2 Experimentální design</b>	<b>25</b>
2.1 Externí validita	25
2.2 Interní validita	30
2.2.1 Randomizace	30
2.2.2 Reaktivita	32
2.3 Multifaktoriální design	38
2.4 Smíšený design	41
2.5 Vícekolové experimenty	45
2.6 Shrnutí	48
<b>3 Výběr participantů</b>	<b>51</b>
3.1 Rekrutace participantů	51
3.2 Složení výzkumného vzorku	52
3.3 Určení velikosti výzkumného vzorku na základě velikosti efektu	58
3.3.1 Chyba prvního typu a hladina významnosti	60
3.3.2 Chyba druhého typu a statistická síla	62
3.3.3 Velikost efektu	63
3.4 Určení velikosti výzkumného vzorku na základě rozpočtu	69
3.5 Shrnutí	70
<b>Literatura</b>	<b>73</b>



Představte si, že vám právě na bankovní účet přišlo 1000 Kč. Odesílatele platby neznáte, ale v poznámce je napsáno, že libovolnou část peněz můžete poslat na účet, jehož číslo je tam také uvedeno. Pokud byste se rozhodovali dle standardní ekonomické teorie, měli byste se rozhodnout jako Homo Economicus a celou částku si nechat pro sebe. Homo Economicus představuje zjednodušený obraz člověka určený pro ekonomické modelování. Tento člověk je neomezeně racionální, sobecký a disponuje neomezenou vůlí (viz Boxík 1: Racionalita). Avšak data z experimentální hry simulující rozhodování v podobných situacích ukazují, že skuteční lidé jsou ochotni vzdát se v průměru až 30 % z obdržené částky ve prospěch druhých. Tento příklad ilustruje zásadní a nenahraditelný přínos experimentů v ekonomii: umožňují ověřovat a rozvíjet teorie postavené na velice zjednodušených představách o lidském chování a rozhodování.

## Boxík 1: Racionalita

Ačkoliv je koncept racionality jedním ze základních stavebních kamenů moderní ekonomie, neexistuje jedna jediná obecně přijímaná definice tohoto pojmu. Zatímco některé definice jsou tak široké, že zahrnují téměř jakékoliv rozhodnutí, některé jsou tak úzké, že vyžadují samostatnou kapitolu na jejich vysvětlení. Pro účely této příručky nám přijde vhodná následující definice – člověk je racionální, pokud *volí nejlepší prostředky k dosažení svých cílů* (Posner, 1997). Například pokud stojíte uprostřed náměstí a máte hlad, porovnáte místní restaurace z pohledu ceny, kvality jídla nebo přívětivosti prostředí. Následně vyberete tu restauraci, která vám přinese nejvíce **užitku** (tedy celkového uspokojení přinocného z porovnání nákladů a přínosů).

Experimentální metody se v ekonomii začaly objevovat na přelomu 50. a 60. let minulého století, a to především v pracích nositele Nobelovy ceny za ekonomii<sup>1</sup> z roku 2002, Vernona L. Smitha. V Evropě pak tuto metodu zkoumání rozvíjel jiný pozdější držitel Nobelovy ceny, německý ekonom Reinhard Selten (i proto má Německo silnou experimentální tradici a může tak vedle Spojených států nebo Nizozemska být vhodnou destinací pro zahraniční stáže). Je však nutné poznamenat, že využívání experimentů bylo v ekonomii spíše kuriozitou. O tom koneckonců svědčí i skutečnost, že samotný termín *experimentální ekonomie* byl poprvé použit

<sup>1</sup> Ve skutečnosti žádná „Nobelova cena za ekonomii“ neexistuje. V tomto oboru je udělována *Cena Švédské národní banky za rozvoj ekonomické vědy na památku Alfreda Nobela* (zkráceně *Nobelova pamětní cena za ekonomii*), kterou – stejně jako u Nobelových cen – uděluje Švédská královská akademie věd. Protože je udělována společně s ostatními Nobelovými cenami, často se toto ocenění označuje pouze jako Nobelova cena za ekonomii.

až v roce 1964<sup>2</sup>. K nechuti ekonomů využívat experimenty přispívala především skutečnost, že tato metoda byla typická pro jinou a na první pohled vzdálenou disciplínu: psychologii.

Když se ekonomie v 18. století vydělila z filosofie (v té době ještě pod názvem *klasická politická ekonomie*), byla její analýza založena především na formální logice. I přesto však Adam Smith (který sám sebe považoval za morálního filosofa, ale obecně je považován za otce ekonomie) v obou svých zásadních knihách *Teorie mravních citů* (1759) a *Pojednání o podstatě a původu bohatství národů* (1776) věnuje značný prostor tématům, která by ještě před několika desítkami let byla označena jako výsostně psychologická. Samotná psychologie pak vznikla o století později a na rozdíl od čím dál tím více matematizované a formalizované ekonomie se zhlédla v experimentálních metodách využívaných především v přírodních vědách. Protože obě disciplíny používaly rozdílné metody, lišil se nejen jejich pohled na způsob formulování nových teorií, ale také oblast jejich zkoumání. Pro psychology byla teorie verbální konstrukt, který reflektuje experimentálně vypořizované pravidelnosti v chování a rozhodování jednotlivců. Pro ekonomy byla teorie souborem matematických nástrojů a teorémů, které umožňovaly analýzu makroekonomických agregátů jako například inflace nebo hrubého domácího produktu (Camerer, 1999). Ekonomy první poloviny 20. století tak individuální rozhodování zajímalo jen okrajově. Oprávněně argumentovali tím, že jejich analýza zahrnuje všechny agenty v dané ekonomice, a proto lze předpokládat *normální rozdělení* (viz Boxík 2: Normální rozdělení) všech vlastností potenciálně ovlivňujících sledovanou veličinu. Na agregátní úrovni by tak byly všechny nadprůměrné projevy chování vyrušeny těmi podprůměrnými a ve výsledku by celkové chování bylo průměrné. Jak již však dnes víme (a jak se dozvíte níže), ne všechny lidské vlastnosti jsou normálně rozděleny a nelze tedy očekávat, že se jejich projevy vždy zprůměrují.

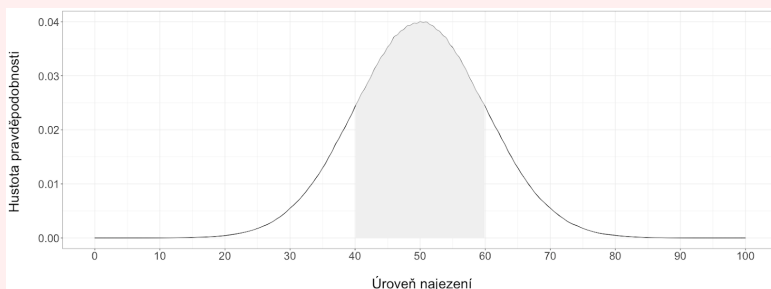
---

<sup>2</sup> Vernon Smith ve své autobiografii vzpomíná, že se na večíрку u přátel v roce 1961 dal do řeči s psychologem a statistikem Sidney Siegelem. V průběhu rozhovoru ke svému překvapení zjistili, že se oba zajímají o „experimentální ekonomii“. Ani jeden z nich toto označení však nepoužil a trvalo další tři roky, než bylo spojení poprvé použito. Sidney Siegel zemřel ve věku 45 let zhruba tři měsíce po tomto večíрку a Smith je přesvědčen, že „*kdyby Sid žil, byl by si nejen vysloužil Nobelovu cenu dříve než my ostatní, ale také by se – možná o mnoho let – urychlilo uznání experimentální ekonomie*“ (Smith, 2018, s. 201). Smith tak uznává, že se mu uznání za průkopnictví tohoto oboru dostalo i proto, že zůstal naživu dostatečně dlouho (Nobelovu cenu obdržel ve věku 75 let).



## Boxík 2: Normální rozdělení

Zkuste o ekonomice nyní přemýšlet jako o velkém obchodu s potravinami, do kterého chodí nakupovat lidé s různými úrovněmi najezení (0–100). Zmapovali jsme úroveň najezení jednoho milionu zákazníků a zjistili jsme, že je normálně rozdělená s průměrem 50 a směrodatnou odchylkou 10. Pokud křivku tohoto rozdělení zaneseme do grafu, kde osa Y vyjadřuje procento lidí s korespondující úrovní energie na ose X, získáme Gaussovu křivku s typickým zvoncovitým tvarem. Jak vidíte na Obrázku 1, křivka je symetrická kolem průměru (v našem případě 50), který je zároveň mediánem a modem (viz Boxík 20: Aritmetický průměr, medián a modus). Normálně rozdělená data se vyznačují tím, že 34,1 % hodnot je vzdáleno maximálně  $\pm 1$  směrodatnou odchylku (SD) od průměru. V obrázku je tato plocha obsahující 68,2 % všech hodnot vybarvena šedivou barvou. Dalších 13,6 % hodnot je vzdáleno  $\pm 1-2$  SD od průměru (oblasti 30–40 a 60–70), 2,1 % hodnot je vzdáleno  $\pm 2-3$  SD od průměru (20–30 a 70–80) a 0,1 % hodnot je vzdáleno více než  $\pm 3$  SD od průměru (0–20 a 80–100). Nyní si představte, že chceme odhadnout celkový příjem obchodu, když víme, že průměrná útrata zákazníků závisí na úrovni jejich najezení: hladoví zákazníci (0–40) utrácí 1200 Kč, optimálně najezení (40–60) utrácí 1000 Kč a přejedení (60–100) utrácí 800 Kč. Protože je úroveň najezení normálně rozdělená, každý zákazník utratí v průměru 1000 Kč. Hladových je totiž stejně jako přejedených, a proto se “zprůměrují”. Po nějaké době se však obchod rozhodne otevřít restauraci, čímž umožní hladovým zákazníkům se před nákupem najíst, a tím se vrchol křivky posune vpravo. Rozdělení již tak není normální, ale zešikmené (skewed) vpravo. V důsledku toho bude průměrná utracená částka nižší než předtím, protože ubude hladových zákazníků, a naopak přibude těch přejedených. Pokud obchod bude nadále předpokládat normální rozdělení úrovně najezení, pak odhady celkového příjmu budou nepřesné.



Obrázek 1: Normální rozdělení

V roce 1944 publikovali Von Neumann a Morgenstern článek *Theory of Games and Economic Behavior*, v jehož příloze formulovali základní principy **Teorie očekávaného užítku** (viz Boxík 3: Teorie očekávaného užítku). Tato teorie, navazující na Bernoulliho **Petrohradský paradox**<sup>3</sup> již z počátku 18. století, umožnila analyzovat rozhodování jednotlivců v situacích zahrnujících riziko (viz Boxík 4: Riziko, nejistota a jednoznačnost).

### Boxík 3: Teorie očekávaného užítku

Teorie očekávaného užítku (Expected utility theory; EUT) umožňuje kvantifikovat, jaký očekávaný užitek (nebo chcete-li uspokojení) přináší rozhodnutí, které s různými pravděpodobnostmi vyústí v různé změny našeho bohatství (Friedman & Savage, 1952). Můžete chtít například zjistit, jaký očekávaný užitek vám přinese jízda hromadnou dopravou na černo. To záleží především na tom, zdali budete odhaleni revizorem. Představte si, že vaše bohatství je 100 tolarů. Dále víte, že pravděpodobnost odhalení je 15 % a v takovém případě musíte zaplatit pokutu ve výši 91 tolarů a vaše bohatství se tak sníží na 9. Druhá možnost je, že s 85% pravděpodobností odhalení nebudete a vaše bohatství zůstane nezměněno. Očekávaný užitek pak vypočítáte jako sumu součinu změny bohatství a pravděpodobnosti, s jakou nastane. V našem případě tedy  $EU = 0,15 \times (100-91) + 0,85 \times (100-0) = 86,4$  utilů. Očekávaný užitek jízdy na černo pak můžeme porovnat s očekávaným užitem jiné alternativy, například jízdy s platnou jízdenkou. Jízdenka stojí 19 tolarů, a pokud si ji koupíte, pak jistě přijedete o 19 tolarů a vaše bohatství se sníží na 81 tolarů. Zároveň jistě nedostane pokutu za černou jízdu a vaše bohatství se jistě nesníží na 9. Očekávaný užitek jízdy s jízdenkou tedy bude  $EU = 1 \times (100-19) + 0 \times (100-91) = 81$  utilů. Pokud nyní porovnáte očekávané užítky obou alternativ zjistíte, že jízda na černo vám přináší více uspokojení ( $86,4 > 81$ ). To však platí pouze pro lidi, kteří mají neutrální vztah k riziku, tedy neodvozují z rizika žádný užitek ani neúžitek. Většina lidí je však k riziku averzní a jejich očekávaný užitek je nižší, pokud pochází z rizikových situací. Základním způsobem, jak tuto lidskou vlastnost ve výpočtu očekávaného užítku zohlednit, je odmocněním rozdílů bohatství. V takovém případě bude očekávaný užitek první alternativy  $EU = 0,15 \times \sqrt{100-91} + 0,85 \times \sqrt{100-0} = 8,95$  utilů a druhé alternativy  $EU = 1 \times \sqrt{100-19} + 0 \times \sqrt{100-91} = 9$  utilů. Vidíte, že při započítání nevole akceptovat riziko se situace otočila a nyní preferujete cestovat hromadnou dopravou s jízdenkou.

<sup>3</sup> Představte si, že si v kasinu můžete zahrát následující hru. Na začátku hry hodíte mincí, a pokud padne panna, obdržíte odměnu 2 eura. Poté hodíte znovu, a pokud padne panna, odměna se zdvojnásobí. Házet můžete až dokud nepadne orel. Hra pak končí a vy získáte odměnu ve výši  $2^{n-1}$  euro, kde  $n$  je počet hodů mincí. Vstup do této hry je však zpoplatněn a otázkou je, kolik peněz jste ochotni zaplatit. Petrohradský paradox (St. Petersburg Paradox) spočívá v tom, že teoreticky byste měli být ochotni za tuto hru zaplatit jakoukoliv částku, kterou po vás bude kasino požadovat. Důvodem je, že očekávaná odměna v každém kole je 1 (2 eura v prvním kole získáte s pravděpodobností  $\frac{1}{2}$ , 4 eura v druhém kole s pravděpodobností  $\frac{1}{4}$ , 8 euro ve třetím kole s pravděpodobností  $\frac{1}{8}$  a tak dále) a očekávaná odměna za celou hru tak je nekonečná ( $1 + 1 + 1 + \dots = \infty$ ). Ve skutečnosti jsou však lidé za takovou hru ochotni zaplatit maximálně nižší desítky eur. Zvažte, že pravděpodobnost výhry 16 eur (tedy 4 po sobě jdoucích panen) je  $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} (\frac{1}{2^4})$ , tedy 6,25 %. Pravděpodobnost výhry 1024 eur je již pouze necelých 0,1 % ( $\frac{1}{2^{10}}$ ).

#### Boxík 4: Riziko, nejistota a nejednoznačnost

V situacích zahrnujících **riziko** (risk) lze objektivně určit pravděpodobnosti jevů, které mohou nastat. Pokud budeme házet s férovou mincí, pak pravděpodobnost, že padne panna je stejná, jako že padne orel, tedy 50 %. Následující dvě situace se liší v tom, že objektivní pravděpodobnost není známá tomu, kdo se rozhoduje (ať již proto, že pravděpodobnost není možné určit, nebo proto, že její zjištění by bylo nákladné).

V situacích zahrnujících **nejistotu** (uncertainty) lze pravděpodobnost jevů, které mohou nastat, určit pouze subjektivně. Představte si, že spolu s kamarády pozorujete na ulici někoho, kdo uzavírá sázky na hod mincí. Za hru musíte zaplatit 100 Kč, a pokud padne panna, vyhraje 200 Kč. Protože si umíte spočítat očekávanou hodnotu této hry (viz poznámka pod čarou číslo 4), je vám jasné, že mince musí být cinklá a panna padá v méně než 50 % případů (kdyby byla mince férová, pak by se takovou loterii nevyplatilo organizovat – výhra by musela být v průměru vyplacena v jednom z dvou hodů). Pár kol hru sledujete a snažíte se určit, jak moc je mince cinklá a jaká je skutečná pravděpodobnost toho, že padne panna. Pokud se o tom budete bavit s kamarády, je možné, že každý z vás dojde k jinému odhadu.

V situacích zahrnujících **nejednoznačnost** (ambiguity) není dostatek informací ani pro subjektivní určení pravděpodobnosti a je proto možné pouze odhadnout rozmezí, ve kterém se skutečná pravděpodobnost bude pohybovat. Vrátime-li se k příkladu se sázkami na hod mincí, pak by se o nejednoznačnost jednalo ve chvíli, kdy nelze pozorovat předchozí hry. Subjektivní pravděpodobnost tak lze odhadnout pouze na základě úvahy typu „aby se vyplatilo loterii provozovat, musí být pravděpodobnost alespoň pod 40 %, nicméně občas musí někdo vyhrát, takže bude vyšší než 10 %“.

(Berger & Bosetti, 2020)

Mezi psychology (ale i některými ekonomy) se samozřejmě našli tací, kteří poukázali na to, že základní předpoklady shrnuté pod pojmem Homo Economicus a ztělesněné v Teorii očekávaného užítku jsou systematicky nesprávné, a tudíž že výsledky modelů založených na těchto předpokladech nemohou být přesné. Jedním z prvních kritiků byl polyhistor a nositel Nobelovy ceny z roku 1978, Herbert Simon, jehož hlavním přínosem do teorie rozhodování bylo prosazování tzv. **omezené racionality** a na ní navazující koncept **uspokojování** (viz Boxík 5: Omezená racionalita a uspokojování).

Ekonomové se v té době však vcelku oprávněně báli, že rozvolnění základních předpokladů a přidávání ad-hoc experimentálních poznatků z nepříliš rozvinuté psychologie by mohlo jejich teorie dovést do slepé uličky. Rozhodli se proto tento rozpor vyřešit následovně. Uznali sice nereálnost teoretických předpokladů, ale zároveň tvrdili, že i přes tuto skutečnost modely na nich založené poskytují relativně přesné výsledky. Tento přístup, pro který se vžilo označení F twist, by tedy šel shrnout asi takto: Jednoduché modely založené na třech předpokladech poskytují uspokojivé výsledky a jakékoliv zpřesnění nestojí za komplikace, které by způsobilo

přidávání dalších předpokladů (Camerer, 1999). Na začátku 50. let 20. století se tak zdálo, že sblížení ekonomie a psychologie byla jednou provždy učiněna přítrž. To do jisté míry potvrdil i jeden z prvních zdokumentovaných experimentů v ekonomii, který provedl mladý francouzský ekonom a taktéž pozdější nositel Nobelovy ceny za ekonomii, Maurice Allais. V roce 1952 byl jedním z organizátorů konference v Paříži, které se účastnili prominentní ekonomové té doby jako například Paul Samuelson, Kenneth Arrow nebo Milton Friedman (po kterém byl pojmenován výše zmíněný F-twist) (Kahneman, 2011). Allais ukázal hostům dvě loterie (1 a 2), které obsahovaly dvě možnosti (A a B). Úkolem hostů bylo z každé loterie vybrat jednu možnost, kterou by preferovali. Tyto loterie vypadaly v principu takto. Než budete číst dále, zkuste si také vybrat.

**Loterie 1A)** 10% pravděpodobnost získat 5000 Kč, 89% pravděpodobnost získat 1000 Kč a 1% pravděpodobnost získat 0 Kč

**Loterie 1B)** 100% pravděpodobnost získat 1000 Kč

**Loterie 2A)** 89% pravděpodobnost získat 0 Kč a 11% pravděpodobnost získat 1000 Kč

**Loterie 2B)** 90% pravděpodobnost získat 0 Kč a 10% pravděpodobnost získat 5000 Kč

Pokud si mohu tipnout, vybrali jste si – stejně jako většina účastníků pařížské konference – možnosti 1B a 2B. Tato kombinace odpovědí však porušuje pravidla racionální volby. V loterii 2 má sice vyšší očekávanou hodnotu<sup>4</sup> možnost 2B (zde jste byli racionální), avšak v loterii 1 má vyšší očekávanou hodnotu možnost 1A, a to především díky 10% šanci získat 5000 Kč. Většina lidí ale preferuje možnost 1B, protože výsledek této možnosti je jistý. Pro tento typ předpokladu nezávislosti se vžil označení **Allaisův paradox** (Allais Paradox). Mechanismus, který způsobuje, že lidé preferují sice méně výhodné, ale zato jisté loterie, byl později nazván **efekt jistoty** (certainty effect). Allais nejspíš očekával, že pokud nejvýznamnějším ekonomům té doby ukáže, že oni sami porušují základní předpoklady vlastních teorií, budou více nakloněni k tomu přehodnotit jak využívání experimentální metody, tak integraci některých psychologických zjištění do ekonomických modelů. Ke svému vlastnímu překvapení zjistil, že tomu tak není. Výsledky jeho experimentu byly označeny za přísloušnou výjimku potvrzující pravidlo a dále ignorovány.

<sup>4</sup> Očekávaná hodnota (Expected Value; EV) může být chápána jako jednodušší forma výpočtu očekávaného užítku pro situace, ve kterých chceme odhadnout očekávanou peněžní hodnotu, a to bez ohledu na roli bohatství a rizikových preferencí. Očekávaná hodnota každé loterie je tedy suma součinu každého prvku a pravděpodobnosti, s jakou nastane. Loterie 1A má  $EV\ 1A = 0,1 \times 5000 + 0,89 \times 1000 + 0,01 \times 0 = 1\ 390$  Kč a loterie 1B má  $EV\ 1B = 1 \times 1000 = 1\ 000$  Kč. Jiný způsob, jakým o očekávané hodnotě můžete přemýšlet, je, že pokud byste loterii 1A hráli 100x, 10x vám přinese 5000 Kč, 89x vám přinese 1000 Kč a 1x vám nepřinese nic. Ze 100 her byste tedy měli 139 000 Kč. Pokud byste 100x hráli loterii 1B, přinesla by vám dohromady 100 000 Kč. Je proto racionální očekávat, že loterie 1A je výhodnější než loterie 1B. Loterie 2A pak má  $EV\ 2A = 0,89 \times 0 + 0,11 \times 1000 = 110$  Kč a loterie 2B má  $EV\ 2B = 0,90 \times 0 + 0,10 \times 5000 = 500$  Kč. Vyberete si tedy 2B, protože očekáváte, že je výhodnější než 2A.

### Boxík 5: Omezená racionalita a uspokojování

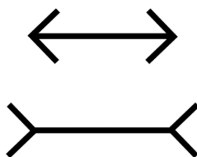
Vraťme se nyní do situace popsané v Boxíku 1. Stále stojíte uprostřed náměstí a chystáte se vybrat si restauraci, ve které se najíte. Protože máte vybitý telefon, musíte restaurace zhodnotit offline. Tedy musíte do každé zajít, podívat se na ceny, zhodnotit přívětivost interiéru a obsluhy, a dokonce i ochutnat pár jídel. I pokud byste byli schopni všechny tyto informace nasbírat a tím k nim získat *přístup* (Access), mohli byste stále mít problém je smysluplně *vyhodnotit* (Assess) nebo se podle nich *zachovat* (Act). Právě tento problém „3A“ ilustruje koncept **omezené racionality** (bounded rationality). Protože lidé nemají neomezený čas a neomezenou kognitivní kapacitu, nelze očekávat, že budou vždy volit nejlepší prostředky k dosažení svých cílů. Proto se většinou uchylují k **uspokojování** (satisficing), tedy k výběru možnosti, kterou na základě omezených informací a času vyhodnotí jako uspokojivou (Simon, 1990). Při rozhodování si pak mohou pomáhat různými přibližnými pravidly. Místo obcházení všech restaurací bychom se tak například mohli jen rozhlédnout a podívat se, jak jsou restaurace zaplněné a vybrat tu, ve které je nejvíce lidí. Naše přibližné pravidlo by spočívalo v předpokladu, že počet lidí v restauraci vypovídá o oblíbenosti daného místa. I přestože naše preference nemusí korespondovat s preferencemi ostatních a teoreticky by bylo možné najít lepší restauraci, využití přibližného pravidla a zvolení uspokojivé možnosti je v dané situaci neefektivnější řešení.

Necelé čtyři roky po této konferenci začal Vernon Smith učit na Purdueově univerzitě (Indiana, USA) základy ekonomie a hledal cestu, jak studentům co nejlépe vysvětlit základní teorii trhů. Využil k tomu experiment simulující obchodování na burze. Polovinu svých studentů dal roli kupujících a druhé polovinu roli prodávajících. Každý student dostal kartičku: kupujícím byla stanovena maximální cena, za kterou mohou aktivum nakoupit a prodávajícím byla stanovena minimální cena, za kterou mohou aktivum prodat. Studenti poté náhodně vytvořili páry kupující-prodávající a jejich úkolem bylo dohodnout se na prodejní ceně. Tomuto mechanismu se říká **dvojitá aukce** <sup>5</sup>. Pokud bylo dohody dosaženo, prodejní cena byla zapsána na tabuli. Pokud nebylo, oba z páru se mohli zkusit dohodnout s někým jiným. Po uplynutí stanovené doby obdrželi studenti nové kartičky a začalo další kolo obchodování. Cílem celého experimentu bylo zjistit, kolik kol potrvá, než trh dosáhne předpovídané rovnovážné ceny a obchodovaného množství. I ke Smithovu překvapení, trh opakovaně dosahoval rovnovážného stavu velice rychle. Mezi lety 1956 a 1960 byl tento experiment s obměnami zopakován a výsledky této série studií byly v roce 1962 publikovány v článku „*An Experimental Study of Competitive Market Behavior*“ (Smith, 1962). O tom, jak složité bylo v této době publikovat experimentální článek v ekonomickém žurnálu svědčí poznámka, že „*publikační proces zahrnoval počáteční odmítnutí, čtyři negativní posudky a dvě revize*“ (Smith, 2018, s. 200).

<sup>5</sup> Na principu dvojitě aukce (double auction) fungují burzy i dnes. „Dohadování“ však probíhá online. Kupující stanoví cenu, za kterou je ochoten nakoupit a prodávající stanoví cenu, za kterou je ochotný prodat. Pokud jsou tyto částky shodné, obchod je uzavřen.

Tato série experimentů přinesla zjištění, které se později stalo základním stavebním kamenem celé metodologie experimentální ekonomie. Smith zjistil, že pokud jsou studenti incentivizováni (finančně motivováni), experimentální trhy dosahují rovnovážného stavu rychleji, než pokud incentivizováni nejsou. V tomto případě by kupující obdržel finanční odměnu ve výši rozdílu rezervační a realizované kupní ceny akcie (tj. pokud by cena akcie byla \$7 a kupující byl ochoten za ni zaplatit maximálně \$10, jeho finanční odměna z této transakce by byla \$3). Podobně finanční odměna prodávajícího by byla ve výši rozdílu mezi rezervační a realizovanou prodejní cenou akcie (tj. pokud by cena akcie byla \$7 a prodávající byl ochoten ji prodat minimálně za \$5, jeho finanční odměna z této transakce by byla \$2).

Zatímco 50. a 60. léta tak šla označit za středověk experimentálních metod, 70. léta můžeme nazývat jejich renesancí. V roce 1976 vydal Smith článek s názvem „*Experimental Economics: Induced Value Theory*“ (Smith, 1976), ve kterém definuje základní metodologii **experimentální ekonomie**. V této dekádě se navíc o výzkum ekonomického rozhodování začali zajímat kognitivní psychologové, především Daniel Kahneman a Amos Tversky, Baruch Fischhoff či Paul Slovic. Jejich východiskem bylo, že podobně jako optické iluze ovlivňují percepci (například Müller-Lyerova iluze, viz Obrázek 2), mohou existovat kognitivní iluze, které ovlivňují rozhodování.



Obrázek 2: Müller-Lyerova optická iluze

*Ačkoliv jsou obě čáry stejně dlouhé, dovnitř směřující šípky způsobují, že spodní čára vypadá delší.*

Ekonomie se ukázala jako vhodná disciplína pro zkoumání kognitivních mechanismů systematicky odklánějících jednotlivce od racionálního chování a rozhodování, protože na rozdíl od psychologie nabízí jasnou definici racionality, proti které lze teorie testovat. Hned na počátku sblížení těchto dvou společenských věd tak došlo k paradoxní situaci: ekonomie byla kritizována za svoji úzkou definici racionality, která však na druhou stranu byla prezentována jako jakýsi ideální stav, kterého by lidé mohli dosáhnout, netrpěli by li různými kognitivními nedostatky. V roce 1974 vydal Tversky s Kahnemanem článek s názvem „*Judgment under uncertainty: Heuristics and biases*“, ve kterém tyto systematické kognitivní nedostatky, vedoucí k chybám

v rozhodování, nazývají **biasy**<sup>6</sup>. Ty jsou produktem nevhodného používání **heuristik** neboli přibližných pravidel, která lidé používají při vyhodnocování situací zahrnujících nejistotu. Tversky a Kahneman ve svém myšlení navazují na Herberta Simona (1990) a koncept omezené racionality a uspokojování. Vycházejí z toho, že není efektivní mít na každý rozhodovací problém dedikovaný kognitivní proces, ale spíše sadu přibližných pravidel, která obsáhnou většinu rozhodovacích problémů, kterým jednotlivec čelí. Čas od času se však může stát, že jednotlivec aplikuje takové přibližné pravidlo, které není určeno pro řešení daného problému. Z pohledu ekonomie je však důležitější druhý článek této dvojice autorů s názvem „*Prospect theory: An analysis of decision under risk*“ (Kahneman, 1979). Prospektová teorie je alternativou k Teorii očekávaného užítku, ve které jednotlivci odvozují užitek z relativní, spíše než z absolutní, změny svého bohatství. Hlavní koncepty spojené s touto teorií jsou referenční bod, vážení pravděpodobností a averze ke ztrátě.

Směr, který se z této tradice zrodil, označujeme jako **behaviorální ekonomii**. Oba termíny, tedy experimentální a behaviorální ekonomie, bývají mnohdy zaměňovány a na první pohled nemusí být patrné, jaký je mezi nimi rozdíl. Experimentální ekonomie je metoda zkoumání v rámci ekonomie, zatímco behaviorální ekonomie je celý myšlenkový směr. Behaviorální ekonomie ke svému zkoumání může využívat experimentální ekonomii, ale zrovna tak může využít například experimentální metody typické pro psychologii. Jak vyplývá z uvedeného přehledu, experimentální a behaviorální ekonomie se vyvíjely paralelně avšak do velké míry odděleně. Došlo proto k paradoxní situaci – zatímco experimentální ekonomie již měla svá jasná pravidla popsána ve Smithově článku z roku 1976, behaviorální ekonomie založená na Prospektové teorii z roku 1979 minimálně jedno z těchto pravidel ignorovala. Tímto pravidlem byly incentivy (tj. variabilní finanční odměna závisící na výkonu participantů ve studii), resp. jejich absence. Prospektová teorie staví na výsledcích několika experimentů (například variace na Allaisův paradox), které jsou založeny na „*odpovědích studentů a zaměstnanců univerzity na hypotetické rozhodovací problémy*“ (Kahneman, 1979, s. 264). Výsledkem bylo, že jedna z nejlivnějších ekonomických teorií druhé poloviny 20. století ignoruje jeden z fundamentů ekonomie – a sice, že drtivá většina rozhodnutí má své náklady a přínosy. Neznamená to nutně, že Prospektová teorie je neplatná, koneckonců i první aukční experimenty Vernona Smitha nebyly incentivizované, a přesto přinesly relevantní výsledky. Avšak tato

<sup>6</sup> Vzhledem k tomu, že experimentální ekonomie je relativně nový obor, který navíc nemá v České republice tradici, veškerá terminologie pochází z angličtiny. Pokud konkrétní termín má ustálený český ekvivalent, dáváme mu přednost a do závorky uvádíme původní anglický termín. Většina termínů však ustálené překlady nemá, pravděpodobně proto, že by překlad zněl zvláštně, nebo by se příliš odchyloval od původního významu. Jedním z takových případů je například právě slovo „bias“, které sice lze do češtiny přeložit jako „sklon k“, nicméně tento překlad plně nevystihuje význam původního termínu. Z toho důvodu v těchto případech používáme původní anglické termíny. Jednou z výhod je, že pokud budete číst odbornou literaturu (což vás jistě nemine, pokud to s experimentální ekonomikou myslíte vážně), budete ihned vědět, o co se jedná.

absence incentív v ekonomickém výzkumu napomohla tomu, že i nadále vznikaly experimenty, u kterých se později ukázalo, že jejich výsledky jsou značně odlišné ve chvíli, kdy jsou jejich účastníci incentivizováni.

V rámci behaviorální ekonomie se kromě rozhodování za podmínek rizika a nejistoty dále začala rozvíjet témata spojená s mentálním účetnictvím, mezičiasovou volbou, teorií her či společenskými preferencemi. Většina z těchto podoborů dnes standardně pro výzkum využívá metod experimentální ekonomie. Tato příručka určená převážně začátečníkům se zabývá metodologií, tedy pravidly, kterými se řídí navrhování experimentů.



# Metodologie experimentální ekonomie

Experimentální ekonomie je kvantitativní metoda (viz Boxík 6: Kvantitativní vs. kvalitativní metoda) založená na sběru dat v experimentálních studiích (nebo zkráceně jen *experimentech*), které simulují různé aspekty lidského chování a rozhodování.

## Boxík 6: Kvantitativní vs. kvalitativní metoda

**Kvantitativní výzkum** je založen především na dotazníkových šetřeních, experimentálních studiích či analýze sekundárních dat. Tento typ výzkumu využívá deduktivní metody, je realizován na větším výzkumném vzorku, získaná data se analyzují většinou statisticky a výsledky je snadnější generalizovat na celou populaci.

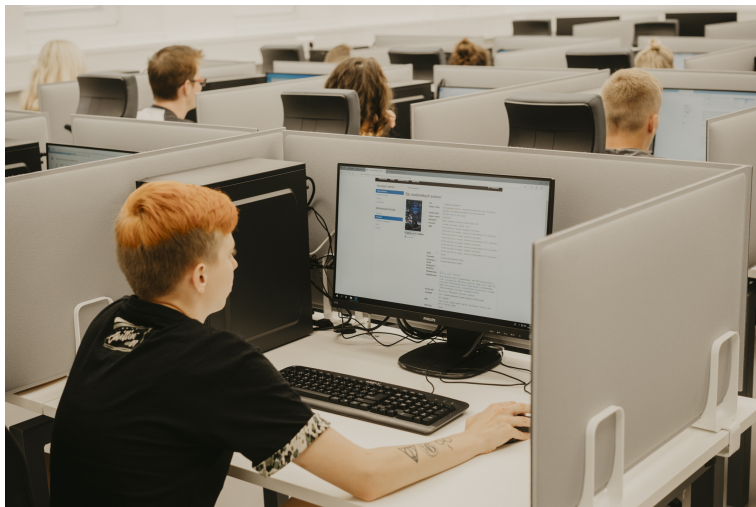
**Kvalitativní výzkum** je založen především na nestandardizovaných osobních rozhovorech, pozorováních či rešerši literatury. Tento typu výzkumu využívá induktivní metody, je realizován na menším výzkumném vzorku, získaná data se analyzují většinou bez využití statistiky a výsledky je obtížné generalizovat na celou populaci.

(Gelo et al., 2009)

Data jsou sbírána v **experimentální laboratoři** – místnosti s počítači, které jsou od sebe odděleny zástěnami tak, aby se **participanti** (účastníci studie) navzájem ovlivňovali co možná nejméně (viz Boxík 7: Participanti vs. subjekty). Kapacity laboratoří jsou typicky mezi 15 a 25 místy, a protože se jedné studie většinou musí zúčastnit minimálně 100 participantů (více o velikosti vzorku naleznete v kapitole 3.3), je třeba data sbírat ve více termínech (kterým se experimentální hantýrkou říká *sessions*). Pokud to povaha experimentu umožňuje, lze data sbírat také online (viz Boxík 8: Online sběr dat).

## Boxík 7: Participanti vs. subjekty

Oba termíny označují účastníky experimentálních studií. Označení „subjekt“ je tradičně spojeno s výzkumem v psychologii nebo v přírodních a lékařských vědách, kde jsou účastníci studií subjektem pozorování. Označení „participant“ se používá především v experimentální ekonomii, kde účastníci nejsou pozorováni, ale na studiích participují.



Obrázek 3: Experimentální laboratoř

Experimenty lze rozdělit na jednoskupinové a víceskupinové. **Jednoskupinové experimenty** většinou slouží k odhadu nějakého parametru, který je později použit k ověření konkrétní teorie či formulaci veřejných politik. Příkladem mohou být experimenty zaměřené na odhad rizikových a časových preferencí nebo na optimální výši zdanění. Do této kategorie by spadaly například i původní experimenty Vernona Smitha popsané výše. Jak vidíte, experimenty mohou, ale nemusí, zahrnovat interakci dvou a více participantů. Právě možnost interakce mezi participanty je jedním z hlavních rozdílů mezi jednoskupinovými experimenty a dotazníky. **Víceskupinové experimenty** pak slouží k určení toho, jak změna zkoumaného parametru ovlivní rozhodování participantů. Participanty jsou experimentálním softwarem náhodně rozděleni minimálně do dvou výzkumných skupin – **kontrolní** (control group) a **experimentální** (treatment group) – jejichž výsledky mezi sebou porovnáваме. Kontrolní skupina typicky obsahuje základní design studie, který navazuje buď na již existující výzkum, nebo teorii. Experimentální skupina pak rozšiřuje základní design právě o faktor, jehož vliv na rozhodování se snažíme prokázat. Příklad nabízí studie (Cueva et al., 2015) zkoumající vliv zvýšené hladiny kortizolu a testosteronu na ochotu podstupovat riziko při obchodování na akciovém trhu. Design kontrolní skupiny byl podobný designu Vernona Smitha a kolegů (Smith et al., 1988). Design dvou experimentálních skupin byl stejný, avšak participantům v první experimentální skupině byla před zahájením studie podána dávka kortizolu a participantům v druhé experimentální skupině byla před zahájením podána dávka testosteronu. Porovnáním výsledků kontrolní skupiny s každou z experimentálních skupin bylo zjištěno, že vyšší úroveň obou hormonů vedly ke

statisticky významnému nárůstu investování do rizikových akcií. Ačkoliv tato studie již spadá do oblasti neuroekonomie, hezky ilustruje koncept kontrolních a experimentálních skupin (viz Boxík 9: Neuroekonomie). Ten totiž nejspíše znáte z prostředí farmakologie, kde pacienti v experimentální skupině obdrží testované léčivo a jejich zdravotní stav je následně porovnáván se stavem pacientů v kontrolní skupině, kteří obdrží placebo. Pro experimentální ekonomii jsou více typické studie, ve kterých se jednotlivé skupiny liší například v poskytnutých informacích, popřípadě v odlišně nastavených incentívách. Víceskupinové studie jsou tak často používány v kontextu konceptů ovlivňujících rozhodování v podmínkách rizika a nejistoty (například různé biasy) nebo vlivu společnosti na rozhodování jednotlivců a skupin, tedy tzv. společenských preferencí.

#### Boxík 8: Online sběr dat

Data do některých studií mohou být sbírána online. Jedná se především o experimenty, které nevyžadují opakované interakce v reálném čase. Online sběr dat probíhá prostřednictvím služeb specializovaných na nábor participantů, jako je například prolific.co. Hlavní výhodou je široká základna participantů (které lze filtrovat na základě různých charakteristik jako pohlaví, věk a tak podobně) a rychlé nasbírání dat (při nižších stovkách participantů lze mít data sebraná v řádu hodin) bez nutnosti disponovat přístupem do experimentální laboratoře. Online sběr dat má však i několik nevýhod. Kromě toho, že experiment musí být dostupný na internetu, je často problematická kvalita dat. Protože jsou lidé za účast placeni, existují „farmy“ participantů, kteří mají za cíl, se co možná nejrychleji proklikat největším počtem studií. Samozřejmě, že při této strategii nevěnují příliš pozornosti tomu, co dělají. Částečným řešením tohoto problému je využívání attention checků, tedy otázek na začátku studie ověřujících porozumění instrukcím. V případě, že participant není schopen tyto otázky zodpovědět, není mu umožněno ve studii pokračovat (stejný způsob kontroly lze v menší míře využívat také v laboratorních experimentech). Dodatečnou nevýhodou pro studie realizované v češtině je, že náborové služby jsou typicky zahraniční a většina registrovaných jsou angličtí mluvčí.

Účelem této části příručky je poskytnout přehled o základních metodologických konceptech, nebo chcete-li pravidlech hry, které je nutné při navrhování vašeho prvního experimentu znát. Pro ilustraci jednotlivých konceptů jsme vybrali několik nejznámějších experimentálních her (Hru na diktátora, Hru na ultimátum, Hru na důvěru a Hru na veřejné statky), které se používají převážně pro výzkum společenských preferencí, avšak jejich modifikace lze nalézt i v mnoha studiích týkajících se jiných témat. V rámci každé hry vždy vysvětlujeme několik souvisejících konceptů. Tato část tak nemá povahu jakéhosi abecedního rejstříku, ale spíše sněhové koule, která na sebe postupně nabaluje další vrstvy experimentálních pravidel.

V první kapitole této části se zabýváme úlohou odměňování participantů v experimentální ekonomii. Dozvíte se, proč je nutné participanty odměňovat za výkon, proč je nesmíte klamat a co je prvotní vybavení. Druhá kapitola je věnována základním principům tvorby experimentálního designu. Naučíte se, jak eliminovat faktory ohrožující externí a interní validitu studie a jak navrhnout experimenty s více výzkumnými skupinami. Dozvíte se, jaký je rozdíl mezi between-subjects a within-subjects designy a jaké jsou specifika vícekolových experimentů. Třetí kapitola poskytuje bližší pohled na účastníky experimentálních studií a naučí vás určit, jak velký výzkumný vzorek k realizaci experimentu potřebujete. Na konci každé kapitoly naleznete její shrnutí.

### Boxík 9: Neuroekonomie

Neuroekonomie hledá vztahy mezi ekonomickým rozhodováním a fyziologickou aktivitou v konkrétních oblastech mozku. Využívá k tomu řadu metod od přímého měření mozkové aktivity (pomocí fMRI, PET nebo EEG), přes chemické (například podáváním hormonů) nebo elektromagnetické (pomocí TMS) ovlivňování funkcí mozku, až po komparativní studie srovnávající rozhodování zdravých pacientů s rozhodováním pacientů s mozkovou dysfunkcí. Možná první takovou studii byl tzv. podivuhodný případ Phinease Gage, mistra na stavbě železnice ve Vermontu. V roce 1948 na stavbě došlo k výbuchu a Gage zasáhla metr dlouhá ocelová tyč o průměru 3 cm – prolétla mu levou tvář a vylétla horní část lebky. Kromě ztráty zraku v levém oku se mu podařilo zotavit natolik, že neměl žádné viditelné následky. V důsledku poškození mozku však došlo k zásadní změně jeho osobnosti, kterou jeho lékař popsal jako „*dětskou v intelektuálních schopnostech a projevech, avšak s animálními pudy silného muže*“. Použitím moderních zobrazovacích technik bylo později určeno, že došlo k poškození části mozku spojené s plánováním budoucnosti, nebo regulací chování (ventromediální prefrontální kortex). K podobnému poškození došlo v důsledku operace nádoru také u pacienta známého jako Elliot. Kromě problémů s plánováním budoucnosti však měl také problémy s rozhodováním projevující se nekonečným zvažováním všech pro a proti. Jedním z dalších projevů jeho stavu byla také patrná absence emocí. Neurovědec Antonio Damasio na základě podobných pozorování navrhl **hypotézu somatických markerů** (somatic marker hypothesis). Somatické markery jsou emoce a s nimi asociované tělesné stavy, které jsou spojeny s konkrétními situacemi a jejich minulými dopady. Slouží tak jako heuristiky, které nám na základě toho, zdali z dané situace máme dobrý, nebo špatný pocit, umožňují rychle učinit rozhodnutí bez vynakládání kognitivního úsilí. Například nemusíte mít žádný vědomý důvod pro oblibu modré barvy, ale protože na vás působí uklidňujícím dojmem, budete ji upřednostňovat před hnědou. Kdybyste podobně jako Elliot nebyli schopni prožívat emoce, nevěděli byste, jakou barvu máte radši. Zdá se tedy, že pro efektivní rozhodování jsou emoce nezbytné (Damasio, 1994). Pokud by vás toto téma zajímalo více, začněte knihou „Kdo to tady řídí?“ od Michaela Gazzanigy.

fMRI = funkční magnetická rezonance

PET = pozitronová emisní tomografie

EEG = elektroencefalografie

TMS = transkraniální magnetická stimulace

Tím hlavním, co odlišuje metodologii experimentální ekonomie od metodologie experimentální psychologie či biologie, je **existence finančních incentív** (odměn) a **absence klamání** (deception). Ačkoliv se to na první pohled nezdá, tyto dva koncepty spolu úzce souvisí. O tom, jak spolu souvisí, pojednává tato kapitola a na příkladu Hry na diktátora představuje další související koncepty.

Ve Hře na diktátora (Dictator game) spolu hrají dvojice účastníků, kterým je náhodně přidělena buď role Diktátora, nebo role Příjemce. Experimentátoři na začátku hry přidělí účastníkům v roli Diktátora **prvotní vybavení** (endowment) ve výši  $e$  ECU (viz Boxík 10: Experimental Currency Unit). Diktátoři jsou instruováni, že mají tuto částku dle svého uvážení rozdělit mezi sebe a Příjemce, se kterým byli anonymně spárováni. Na základě tohoto rozdělení pak bude oběma účastníkům vyplacena finanční odměna za účast v experimentu. V této hře Příjemce nemá na výši své odměny žádný vliv a je tak ponechán napospas Diktátorovi. Mimochodem, právě tuto hru jsme použili na samém začátku Úvodu do této příručky. Diktátoři typicky nabízejí Příjemcům v průměru 20–30 % svého prvotního vybavení (Camerer, 1997). 36 % Diktátorů nenabídne Příjemcům nic, 17 % Diktátorů se rozhodne rozdělit částku rovným dílem a 5 % Diktátorů se rozhodne dát Příjemcům celou částku (Engel, 2011). Tato jednoduchá experimentální hra z roku 1994 ukazuje, že většina lidí se nechová jako Homo Economicus, který by si měl vždy nechat celou částku (Forsythe et al., 1994).

## 1.1 Úloha incentív

V experimentálních studiích téměř ve všech oborech bývá zvykem platit účastníkům studie **show-up fee**, neboli fixní odměnu za účast. Tato odměna je předem stanovená a není závislá na „výkonu“ účastníků v dané studii<sup>1</sup>. Referenčním bodem pro stanovení výše fixní odměny je výše průměrné hodinové mzdy za brigádu. Jelikož účast v laboratorních experimentech musí být dobrovolná, zapojují se do nich pouze lidé, kteří to z nějakého důvodu považují za zajímavé. Existence finanční odměny pomáhá do té míry, že se experimentů neúčastní pouze lidé se zájmem o vědu. Jejich motivace pro účast může být totiž velice odlišná od motivace běžných lidí zapojujících se do každodenních mezilidských a tržních interakcí.

<sup>1</sup> Jedním z důvodů pro existenci show-up fee je fenomén **samovýběru** (self-selection). V případě absence jakékoliv finanční odměny za účast by se laboratorních experimentů účastnili převážně lidé se zájmem o vědu. To je však velice specifická část populace, jejíž rozhodování ve studii může být ovlivněno jiným motivem než rozhodování „běžných“ lidí zapojujících se do každodenních tržních interakcí. Využití vzorku dobrovolníků by proto ohrozilo externí validitu studie.

### Boxík 10: Experimental Currency Unit

V experimentálních studiích se často místo skutečné měny používá experimentální měna často označovaná ECU (Experimental Currency Unit), nebo MU (Monetary Unit). V instrukcích na začátku experimentu je pak uveden konverzní poměr, kterým se ECU převádí do lokální měny. Například Diktátoři tak nerozdělují 250 CZK, ale 250 ECU s konverzním poměrem 1 ECU = 1 CZK. Říkáte si, proč zadání experimentu komplikovat, když je nakonec konverzní poměr 1:1? Využívání ECU má několik výhod. První výhoda souvisí s replikabilitou studie. Pokud by tedy byl experiment kromě Česka realizován ještě například v zemi platící eurem, stačilo by pouze změnit konverzní poměr na 25 ECU = 1 EUR. Lze si také představit situaci, ve které chceme replikovat experiment, který byl původně pořádán ve Spojených státech, kde prvotní vybavení bylo 10 USD. Kdybychom chtěli porovnat výsledky z České republiky s těmi ze Spojených států, bylo by dobré zachovat přerozdělování 10 jednotek (původně dolarů), a tím minimalizovat možný efekt přerozdělování jiné nominální částky. Lze proto 10 USD upravit na 10 ECU a přidat informaci o konverzním poměru do korun. Druhá výhoda je, že ECU lze v případě potřeby využít spíše jako body než jako měnu. Výše jsme zmínili možnost vyplácení participantů formou turnaje, ve kterém stanovený počet participantů s nejvíce ECU získá stanovenou odměnu. Bylo by matoucí, kdyby instrukce obsahovaly větu „Participant s nejvyšším získaným počtem CZK získají odměnu ve výši 200 CZK.“. Třetí výhoda souvisí s přípravou experimentálního softwaru, kdy se u složitějších studií objevují různé částky na několika místech. Při používání ECU nehrozí, že se například při změně prvotního vybavení zapomenou v softwaru upravit nějaké údaje. Stačí upravit konverzní poměr v instrukcích.

Při přemýšlení o výši show-up fee však musíme vzít v úvahu ještě další okolnosti. Jednou z nich je druhá složka odměny, typická pro ekonomické experimentální studie. Tato variabilní část odměny, typicky závislá na výkonu participanta ve studii, se v experimentální ekonomii označuje pojmem **incentivy**. Průměrná výše odměny za výkon by přitom neměla být nižší než výše odměny za účast. Toto pravidlo souvisí s úlohou incentív v experimentální ekonomii, kterou lze rozdělit na praktickou a teoretickou.

Praktická úloha incentív spočívá v tom, že více motivuje participanty k tomu, aby se ve studii snažili a byli pečliví. Složitější experimenty obsahují detailní instrukce o tom, v čem studie spočívá a jak funguje. Pokud participant obdrží pouze odměnu za účast, nemají žádnou dodatečnou motivaci instrukce pečlivě číst a poté se podle nich chovat. Z ekonomického pohledu by šlo dokonce říci, že racionální participant budou v experimentu dělat jen nezbytné minimum pro to, aby se „proklíkali“ až na jeho konec, protože jsou placeni pouze za účast. Nikoliv však za aktivní účast, která jim nepřináší žádný dodatečný zisk. Někdo může namítnout, že lidé přece věci nedělají jen pro peníze. To je samozřejmě pravda, nicméně téměř veškerá rozhodování, která v životě děláme, mají náklady a přínosy. Zejména z praktických důvodů pak většinu nákladů a přínosů v ekonomických experimentech aproximujeme penězi. Je totiž

obtížné určit, jak různí participanti vnímají náklady na rozhodnutí (například čas a úsilí). Je však jednoduché konstatovat, že pokud rozhodnutí bylo učiněno, přínos v podobě peněz převážil nad těmito náklady.

Teoretická úloha incentív navazuje na tu praktickou. Incentivy jsou typicky nastaveny tak, aby motivovaly participanty k chování, které předepisuje testovaná teorie. V případě Hry na diktátora jsou Diktátoři incentivizováni k tomu, aby se chovali jako Homo Economicus a nechali si celé prvotní vybavení pro sebe a Příjemci neposlali nic. Designy těchto experimentů – ve kterých se participantům vždy vyplatí chovat se podle svých skutečných preferencí – se označují jako kompatibilní s incentivy (incentive compatible) (Hurwicz, 1972; Smith, 1982). Pokud předpokládáme, že participanti vždy upřednostňují vyšší odměnu před nižší odměnou, pak výsledky odchylovající se od teoretických předpovědí naznačují, že testovaná teorie je neúplná. Pokud například většina Diktátorů posílá Příjemcům nenulové částky, znamená to, že jejich rozhodování je ovlivněno něčím, co teorie nepředpokládá. Fehr-Schmidtův model (1999) averze k nerovnosti navrhuje, že o rozhodování Diktátorů rozhoduje vztah mezi jejich averzí k výhodné nerovnosti (nerovnost, ze které Diktátor benefituje) a averzí k nevýhodné nerovnosti (nerovnost, která znevýhodňuje Příjemce). Tato teorie tak umí vysvětlit, proč většina Diktátorů posílá Příjemcům větší než nenulové, ale menší než poloviční částky.

Nyní si představte, že bychom hráli Hru na diktátora bez incentív. To znamená, že participanti by sice dostali odměnu za účast, ale obdržené prvotní vybavení by rozdělovali pouze teoreticky – ani oni, ani Příjemci by neobdrželi žádné další peníze. Největší problém je, že v této situaci jako výzkumníci nemáme ponětí, co participanti v roli Diktátorů maximalizují. Možná se opravdu snaží vžít do situace, ve které by rozdělovali skutečné peníze a chovají se, jako kdyby byl experiment incentivizovaný. Možná se však jen chtějí cítit dobře (minimalizovat pocívanou újmu z jimi způsobené nevýhodné nerovnosti) a pošlou Příjemci mnohem více, než kdyby rozdělovali „vlastní“ peníze. Možná se tak chovají proto, že chtějí potěšit experimentátory v domněnání, že ti testují, jak jsou na sebe lidé hodní. A možná také nepošlou vůbec nic, protože si uvědomují, že je to celé jen teoretické cvičení a nikoho tím neovlivní. Nejpravděpodobnější však je, že odpovědi v neincentivizované Hře na diktátora by byly motivovány kombinací podobných úvah. Abychom se v experimentální ekonomii vyvarovali této interpretační nejistotě, rozhodování musí být vždy incentivizováno. Tento prvek ekonomických experimentů je tak zásadní, že je podmínkou pro publikování výsledků v ekonomických žurnálech.

Relativně často zmiňovaným argumentem proti využívání incentív je, že ekonomické experimenty jsou kvůli nim finančně náročné, což může být problém pro začínající výzkumníky, kteří ještě nedisponují dostatečně velkými rozpočty. Zde je však potřeba uvědomit si, že většina výzkumníků získává první zkušenosti s experimenty ještě během svého magisterského (někdy i bakalářského) studia – buď jako součást týmu seniornějších výzkumníků, nebo pod vedením svého mentora. Otázky rozpočtu studie tak neřeší, nebo jim se získáním rozpočtu pomáhají zkušenější kolegové. Někdy pak lze také uvažovat o vyplácení participantů formou turnaje. V případě diskutované Hry na diktátora lze například říci, že prvních  $X$  Diktátorů s nejvyšším počtem ECU získá předem stanovenou finanční odměnu. Problémem v tomto případě je, že tato změna vyplácecího mechanismu dává Diktátorům extra motivaci k tomu, nechat si co nejvíce. Odměňování turnajem tak může být vhodnější například pro Hru na ultimátum, ve které má Příjemce možnost nabídku odmítnout (více se o této hře dozvíte v kapitole 3.2). Jak vidíte, při navrhování experimentů je třeba věnovat pozornost každému aspektu studie a neustále si pokládat otázku, jak určitá změna v designu může ovlivnit chování a rozhodování participantů.

## 1.2 Zákaz klamání participantů

Druhou podmínkou pro publikování výsledků ekonomických experimentů je absence klamání participantů (deception). Za klamání se považuje nejen lhaní, ale také zatajování důležitých okolností experimentálního designu – především těch, které se přímo týkají mechanismů určujících výši konečné odměny. Instrukce k experimentu musí obsahovat všechny informace, na základě kterých participanté poté dělají svá rozhodnutí, ale zároveň by měly být co nejstručnější. Někdy tak stojíte před dilematem, zdali nějaký konkrétní detail studie máte do instrukcí zahrnout, nebo ho v rámci stručnosti vynechat. Například u Hry na diktátora musíte participantům říci, že jim náhodně bude přidělena role Diktátora, nebo Příjemce, že Diktátor bude rozdělovat částku  $e$  ECU a že Příjemce toto rozdělení musí přijmout. Technické detaily o tom, jakým způsobem participanty do rolí randomizujete, již vysvětlovat nemusíte. Ve složitějších designech však toto rozlišení toho, co je a není důležité, nemusí být tak jednoznačné. Hlavním vodítkem pak je odpověď na otázku, zdali vynechání nějaké informace může ovlivnit rozhodování participantů – pokud ano, je třeba informaci zahrnout. V případě, že si nejste jistí, je informaci také lepší zahrnout a neriskovat tak případné klamání participantů.

Důvodem pro toto striktní pravidlo je potřeba zachování důvěry participantů v pravdivost poskytnutých informací. Chceme si být totiž co možná nejvíce jistí, že hlavní motiv, který participanté sledují, je maximalizace finanční odměny. Problém klamání je, že jakmile jím jednou důvěru participantů narušíte, při jejich příští účasti si nemůžete být jistí, zdali vám věří, nebo ne. Představte si například situaci,



kdy participanty necháte hrát Hru na diktátora s tím, že si peníze, se kterými hrají, mohou ponechat. Na konci studie byste jim však řekli, že celá hra byla pouze hypotetická a že žádné peníze nedostanou. Nejspíše byste participanty našťvali tak, že by už nikdy nemuseli přijít. I pokud by přišli, nevěděli byste, podle čeho se rozhodují – podle toho, co jim říkáte, nebo podle toho, co si myslí, že jim říkáte? V příští hře (ve které byste jim peníze opravdu dali) by pak například mohli být Diktátoři velice štědrí. Důvodem by však nebyly jejich preference, avšak domněnka, že stejně žádné peníze nedostanou.

Klamáním participantů ve své studii byste navíc poškodili i laboratoř, ve které sběr dat pořádáte a tím i výzkum všech kolegů, kteří laboratoř také využívají. Participantů totiž málokdy rozlišují, kdo pořádá jakou studii a svou účast mají spojenou s laboratoří samotnou. Z toho důvodu je také běžné, že se ve stejné laboratoři nepořádají zároveň ekonomické experimenty a experimenty jiných oborů (například psychologie), které jsou někdy nuceny – vzhledem k absenci incentív – využívat klamání. Někteří výzkumníci dokonce tvrdí (s trochou nadsázky), že laboratoř pro ekonomické experimenty by neměla být ani ve stejné budově jako ta pro experimenty jiných oborů.

### 1.3 Prvotní vybavení

Prvotní vybavení (endowment) jsou peníze (nebo jejich ekvivalent v ECU), které participantů obdrží na začátku studie a během studie s nimi disponují. V případě Hry na diktátora je to částka  $e$ , kterou Diktátoři na začátku hry obdrží k rozdělení. Prvotní vybavení je tedy základ variabilní odměny participantů. Skutečnost, že participantů prvotní vybavení obdrží však (spolu se skutečností, že většina našich rozhodnutí se týká peněz, které jsme si nějak zasloužili) vytváří hrozbu pro externí validitu. Pro nenadálý jednorázový příjem se v ekonomii používá označení **windfall money**. Pojem **house money efekt** pak označuje situaci, kdy jsou lidé ochotni s penězi „navíc“ více riskovat (Cárdenas et al., 2014). Jedním z vysvětlení pro oba koncepty je skutečnost, že lidé za obdrženými nebo vyhranými penězi nevidí hodnotu, která je u běžného příjmu generována prací. U Hry na diktátora bylo například ukázáno, že Diktátoři posílají Příjemcům výrazně nižší částky, pokud si své prvotní vybavení musí nejdříve zasloužit (Cherry et al., 2002). Pokud tedy nemáme dobrý důvod participantům prvotní vybavení darovat, musíme je nechat si ho zasloužit. To je možné udělat minimálně dvěma způsoby – vyplněním dotazníku nebo splněním úkolu.

## Vyplnění dotazníku

Dotazníkem pro tyto účely rozumíme již publikovanou baterii otázek, jejíž vyhodnocení poskytne ideálně jedno skóre, které můžeme zahrnout do analýzy. Tím zaprvé získáme dodatečnou informaci o participantech a zadruhé jim umožníme zasloužit si prvotní vybavení, které později využijí během experimentu. Využití dotazníku však má několik omezení, která byste při jeho použití měli zvážit. Vyplnění dotazníku by nemělo trvat desítky minut – nezapomeňte, že participantům platíte show-up fee (které může být odvozeno od počtu hodin strávených v laboratoři). Krátký dotazník také zajistí, že participanti nebudou unavení nebo se nebudou nudit ještě předtím, než se dostanou k hlavní části experimentu. Dotazník by také neměl prozrazovat, na co se výzkum zaměřuje – pokud dotazník nevyužíváte pro vygenerování prvotního vybavení, pak je dobrou praxí ho umístit až za hlavní část experimentu. V neposlední řadě bych pak začátečníky důrazně odradil od vytváření vlastních dotazníkových otázek – výstup z vlastního dotazníku nebude jedno skóre, ale řada jednotlivých odpovědí, které poté musíte analyzovat. Může se navíc stát, že některé z otázek nedopatřením naruší validitu celého experimentu nebo způsobí neudělení souhlasu etické komise (viz Boxík 11: Etická komise).

### Boxík 11: Etická komise

V případě experimentálního výzkumu počítejte s tím, že váš výzkumný záměr bude muset být schválen etickou komisí vaší instituce (typicky fakulty nebo univerzity). Pokud je však výzkum anonymní, neinvazivní a účastní se ho dospělí lidé, jeho schválení by mělo být pouhou formalitou.

## Splnění úkolu

Druhou možností je nechat participanty splnit úkol (effort task), a tím si zasloužit své prvotní vybavení<sup>2</sup>. V tomto kontextu hledáme jednoduchý úkol, který bude každý participant schopen splnit – potřebujeme totiž, aby všichni Diktátoři rozdělovali stejnou sumu peněz. Jedním z vhodných úkolů je například „Počítání nul“ – v tomto úkolu je participantům zobrazena tabulka složená pouze z jedniček a nul a jejich úkolem je spočítat, kolik nul tabulka obsahuje. Na zvážení poté je, zdali necháme

<sup>2</sup> Pokud potřebujete určit úsilí participantů v jiných částech experimentu (tj. ne pro vygenerování prvotního úsilí), můžete využít takzvané deklarované úsilí (stated effort). Při využití této metody participantů deklarují úroveň úsilí (typicky od 0 % do 100 %), kterou chtějí vyvinout. Každá úroveň deklarovaného úsilí je pak zpoplatněna určeným počtem ECU. Například ve Hře na výměnu dárků (Gift exchange game) spolu hrají Zaměstnavatelé a Pracovníci. Zaměstnavatelé nabízí Pracovníkům kontrakt, který obsahuje závaznou úroveň mzdy ( $w$ ; 0–100 ECU) a nezávaznou úroveň požadovaného úsilí ( $e$ ; 0–100 %), kterou Pracovníci mohou (ale nemusí) vyvinout. Odměna Pracovníků pak je vypočítána jako  $w - c(e)$ , kde  $c(e)$  je rostoucí funkce deklarovaného úrovně úsilí. Jinými slovy, odměna Pracovníků je tím menší, čím vyšší úroveň úsilí se rozhodnou vyvinout. Úsilí vyvinuté nad požadovanou úroveň je „dar“ Pracovníků svým Zaměstnavatelům.

participanty dělat tento úkol pouze jednou, nebo jim například dáme časový limit, ve kterém musí vyřešit co nejvíce tabulek s tím, že za to dostanou své prvotní vybavení. Podobně jako u dotazníku by i tyto úkoly měly trvat maximálně jednotky minut, abychom participanty příliš nevyčerpali. Následující Tabulka 1 uvádí některé příklady úkolů, jejichž širší přehled naleznete v článku Charness et al. (2018). Tabulka mimo jiné uvádí, jak jsou jednotlivé úkoly náročné na přípravu (sloupeček *Náročnost přípravy*), zda se účastníci v průběhu plnění mohou zlepšovat učním (*Prostor pro učení*) a zda se výsledky při plnění úkolu budou lišit dle dovedností účastníků (*Vliv dovedností*).

Název	Popis	Náročnost přípravy	Prostor pro učení	Vliv dovedností
Zadávání dat z knihovny	Účastníci zadávají informace ze stohu knih do počítačové databáze.	velká	malý	malý
Počítání nul	Na monitoru je náhodně rozmístěno několik nul a jedniček. Participanti počítají, kolik nul na monitoru je.	malá	malý	malý
Řešení bludiště	Participantů procházejí počítačovým bludištěm pomocí šipek na klávesnici.	malá	malý	střední
Tetris	Participantů hrají 4 řádky Tetrisu.	malá	malý	malý
Třídění a počítání mincí	Participantů musí za určitou dobu roztrždit a spočítat určitý počet mincí dle jejich nominální hodnoty.	střední	střední	malý
Přepis řeckých písmen	Na obrazovce se objeví řada náhodných a rozmazaných řeckých písmen. Participant musí najít a kliknout na odpovídající písmeno na seznamu řeckých písmen.	malá	malý	malý
Nemožný matematický problém	Účastníci obdrží sadu čísel a je jim řečeno, že musí vybrat skupinu čísel, jejichž součet je 100. Úlohu však nelze vyřešit, protože žádná kombinace čísel v součtu není 100. Úsilí se měří jako čas strávený nad úkolem, než se účastníci vzdají.	malá	velký	střední

Tabulka 1: Přehled úkolů k zasloužení prvotního vybavení

Zdroj: Charness et al. (2018)

Možná se teď ptáte, na základě čeho se máte rozhodnout, zdali pro vygenerování prvotního vybavení použít dotazník, úkol, nebo úplně něco jiného. Podobných otázek zřejmě budete mít během čtení této příručky mnoho a bohužel na všechny existuje pouze jedna odpověď – „to záleží“. Záleží na tom, na jakou (výzkumnou) otázku se experimentem snažíte odpovědět. Zvažte následující tři výzkumné otázky:

1) *Jaká je výše příspěvku ve Hře na diktátora, pokud je prvotní vybavení dar?*

V tomto případě je darování prvotního vybavení součástí experimentu, a proto ho není třeba nikterak generovat.

2) *Jaká je výše příspěvku ve Hře na diktátora, pokud je prvotní vybavení zasloužené?*

Nejjednodušší cestou, jak v tomto typu experimentů vygenerovat prvotní vybavení, je využití úkolu, například zmíněného hledání nul. Nespornou výhodou podobných úkolů je skutečnost, že minimalizují riziko jakéhokoliv ovlivnění pozdějšího rozhodování participantů (na rozdíl od dotazníků, viz níže). Na druhou stranu mohou podobné úkoly působit trochu uměle – participantům bude zřejmě jasné, že tento úkol slouží jen k vygenerování prvotního vybavení. Dobrou zprávou je, že co se týká úkolů, můžete být poměrně kreativní a vymyslet si vlastní, který se bude tvářit „užitečně“ (pokud však teprve sbíráte zkušenosti s navrhováním experimentů, doporučuji kreativitu omezit a držet se postupů, které již byly publikovány).

3) *Jaký vliv má prosocialita na výši příspěvků ve Hře na diktátora, pokud je prvotní vybavení zasloužené?*

K zodpovězení této otázky potřebujeme kromě samotné hry také nějakým způsobem kvantifikovat prosocialitu participantů. K tomu je vhodné využít dotazník, například *Prosocialness Scale for Adults (PSA)*. Musíte však myslet na to, zdali výši příspěvků nemůže ovlivnit skutečnost, že se nejdříve ptáte na otázky spojené s prosocialitou. To by ze strany participantů mohlo být vědomé (participantů se budou chtít vyhnout kognitivní disonanci (viz Boxík 12: Kognitivní disonance) a příspěvky budou korespondovat s jejich odpověďmi v dotazníku), nebo i nevědomé (pokládání otázek zaměřených na prosocialitu by participanty „přepnulo“ do více společenského módu). Nejlepší recept na vyhnutí se těmto problémům je umístit dotazníky za hlavní část experimentu. To samozřejmě udělat můžete a k vygenerování prvotního vybavení využít úkol. Tím však zaprvé experiment prodloužíte (i když v tomto případě to není nikterak zásadní) a zadruhé i rozhodnutí o výši příspěvku může ovlivnit odpovědi v dotazníku, který bude následovat (například participantů, kteří se rozhodnout neposlat Příjemci nic, se poté v dotazníku mohou vyhýbat prosociálním odpověďmi).

Jak můžete vidět, často se při navrhování experimentu dostanete do situace, ve které každé řešení má nějaké úskalí. Vy se pak musíte rozhodnout, které z možných ovlivnění je, vzhledem k vašemu záměru, nejpříjemnější. Bohužel, ne na všechny problémy, se kterými se můžete setkat, existují jasná řešení.

#### Boxík 12: Kognitivní disonance

Kognitivní disonance je pojem označující pocítovaný vnitřní konflikt nastávající v situacích, ve kterých se člověk chová v rozporu se svou představou o tom, jaký by chtěl být, nebo jak by chtěl být veřejně vnímán. Lidé se snaží tomuto napětí vyhnout přizpůsobením postojů svému chování spíše než změnou chování samotného. Příkladem je situace, ve které si lidé odůvodňují nechození k volbám myšlenkou, že jeden hlas stejně nic nezmění.

(Dickerson et al., 1992)

## 1.4 Shrnutí

### Experimentální ekonomie

Experimentální ekonomie je kvantitativní metoda založená na sběru dat v experimentálních studiích, které mohou být buď jednoskupinové, nebo víceskupinové. Jednoskupinové experimenty typicky slouží k odhadu parametrů (například výše příspěvku ve Hře na diktátora). Víceskupinové experimenty slouží k určení toho, jak změna zkoumaného parametru ovlivní rozhodování participantů (například jak původ prvotního vybavení ovlivní výši příspěvku ve Hře na diktátora). Pokud studie porovnává vliv přítomnosti a absence zkoumaného faktoru, pak jsou participanté rozděleni do kontrolní skupiny (bez zkoumaného faktoru) a do jedné, nebo více experimentálních skupin (které se od sebe liší úrovněmi zkoumaného faktoru, nebo více faktory). Příkladem by byla například studie zkoumající, zdali má administrace kortizolu a testosteronu vliv na objem obchodování na trhu s aktivy.

### Fixní a variabilní odměna

Ekonomické experimenty se vyznačují tím, že kromě fixní odměny za účast (show-up fee) využívají také variabilní složku odměny závislou na výkonu participantů (variabilní odměně se často říká incentive). Protože ekonomická teorie předpokládá, že vyšší získaná odměna je lepší než nižší získaná odměna, rozhodnutí vedoucí k nižším odměnám indikují, že roli hrají ještě jiné faktory. Dobrým příkladem je Hra na diktátora, ve které Standardní ekonomie předpokládá, že si Diktátoři ponechají celé prvotní vybavení. Experimentální data však ukazují, že

Diktátoři sdílejí v průměru 20–30 %. Jedním z možných důvodů pro takové chování je averze k nerovnosti: Diktátoři jsou ochotni vzdát se části odměny proto, aby se vyhnuli špatnému pocitu pramenícího z nerovnosti, kterou by ponecháním si celé částky vytvořili. Experimenty kompatibilní s incentivy také redukuje interpretační nejistotu způsobenou tím, že výzkumníci neví, co motivuje rozhodování participantů. Výši fixní odměny ovlivňuje více faktorů, avšak typicky se pohybuje na úrovni průměrné hodinové mzdy za brigádu. Variabilní odměna by pak neměla být nižší než fixní.

## Absence klamání a důvěra participantů

Kromě nutnosti incentivizovat rozhodování participantů je také naprosto nezbytné vyhnout se jejich klamání (deception). Za klamání se považuje lhaní a úmyslné, či neúmyslné zatajování důležitých okolností experimentálního designu. Za obzvlášť závažný prohřešek se pak považuje klamání v oblasti informací týkajících se výše odměny a jejího určení: participantů musí přesně vědět, jak jejich rozhodování ovlivňuje získanou odměnu. Vodítkem při rozhodování o tom, co vše do experimentálních instrukcí zahrnout, je odpověď na otázku, zdali vynechání nějaké informace může ovlivnit rozhodování participantů – pokud ano, je třeba informaci zahrnout. Klamání je zakázáno především kvůli důvěře participantů v pravdivost poskytnutých informací a pokud byste se ho dopustili, výsledky vašeho výzkumu nebudou publikovatelné. Klamáním ve své studii byste navíc poškodili i samotnou laboratoř a potenciálně tak i výzkum ostatních kolegů, protože participantů většinou nerozlišují mezi jednotlivými studiemi pořádanými ve stejné experimentální laboratoři.

## Prvotní vybavení

Prvotní vybavení jsou peníze, které participantů obdrží na začátku studie a se kterými následně během experimentu disponují. V závislosti na typu studie, respektive na předmětu jejího zkoumání, může být záhodno, aby si participantů prvotní vybavení zasloužili. Pokud částku pouze obdrží, mohou ji vnímat jako windfall money a rozhodovat se o ní odlišně. Příkladem je house-money efekt popisující zvýšenou ochotu riskovat, pokud se lidé rozhodují o penězích, které předtím vyhráli. Existují dvě možnosti, jak si prvotní vybavení zasloužit: vyplněním dotazníku, nebo splněním úkolu. Pokud participantů necháte nejdříve vyplnit dotazník, využijte běžně používané sady otázek (například zkrácenou verzi Big Five) a dejte si pozor, aby z nich nešlo usuzovat na váš výzkumný záměr. Druhou možností je splnění jednoduchého úkolu (například hledání a počítání nul v tabulce jedniček a nul), který bude každý participant schopen splnit. Vyplnění dotazníku, ani splnění úkolu by participantům nemělo zabrat více než vyšší jednotky minut.

Při čtení o prvním experimentálním trhu Vernona Smitha vás možná napadlo, že studenti snažící se dohodnout na ceně jednoho imaginárního aktiva mají daleko do skutečných makléřů obchodujících na Wall Street. Možným nesouladem mezi skutečným světem a výsledky laboratorního experimentu se zabývají koncepty **externí a interní validity** (kapitola 2.1 a 2.2). Osvojení si těchto konceptů obsahujících základní „dos and don'ts“, je základním předpokladem pro navržení dobrého experimentálního designu. Kapitoly 2.3 a 2.4 se věnují multifaktoriálním designům, které typicky obsahují několik výzkumných skupin. V závislosti na tom, zdali se skupin účastní jiní nebo stejní participanti, se tyto studie nazývají **between subjects** nebo **within-subjects design**. Poslední kapitola 2.5 pak diskutuje specifika studií, ve kterých se participanti účastní více experimentálních kol.

## 2.1 Externí validita

Externí validita je *míra, s jakou lze výsledky laboratorního výzkumu použít ve skutečném světě* (Roe & Just, 2009). Je jasné, že design laboratorního experimentu nikdy nebude (a ani nemá ambici být) natolik komplexní, aby zachytil kontext a všechny aspekty zkoumaného rozhodování. V experimentálním designu jde vlastně o přesný opak – cílem je vytvořit design tak jednoduchý, aby nám umožnil zkoumaný aspekt izolovat a kvantifikovat jeho význam.

Při navrhování experimentu je tedy používána **abstrakce** – proces odstraňování nebo zobecňování detailů zkoumaného konceptu za účelem zaměření pozornosti na aspekty, kterým je výzkumníky přikládána větší důležitost (Colburn & Shute, 2007; Kramer, 2007). Standardní ekonomie je disciplína založená na modelování, které je ve své podstatě to samé, co abstrakce. Vzpomeňte si například na výše zmíněný F-twist. Ten ospravedlňoval model Homo economicus tím, že navzdory nereálnosti předpokladů poskytuje relativně přesné předpovědi. Jinými slovy, i přes vysokou míru abstrakce tento model vykazoval vysokou externí validitu. Všimněte si, že míra abstrakce jde ruku v ruce (je korelovaná) s externí validitou. O tomto vztahu můžete přemýšlet tím způsobem, že čím více se bude model (experiment) množstvím svých parametrů blížit realitě, tím méně zobecnitelné jeho výsledky budou. Například pokud bychom v laboratoři vytvořili přesnou kopii burzy na Wall Street a experimentu by se účastnili makléři z Wall Street, výsledky takového výzkumu by byly zobecnitelné pouze na burzu v New Yorku, ale již ne ve Frankfurtu nebo Londýně. Netřeba dodávat, že vytvoření přesné kopie je nereálné, a i kdyby bylo, jednalo by se zřejmě o jeden z nejdražších výzkumů na

světě. Tím narážíme na další výhodu abstrakce – čím jednodušší experimentální design je, tím méně nákladný je<sup>1</sup>. I přestože vyšší míra abstrakce je obecně lepší, výzkumníci si při navrhování experimentu musí dávat pozor, aby to – lidově řečeno – nepřehnali. Pokud by se stalo, že během abstrakce dojde ke zjednodušení, nebo dokonce úplnému odstranění mechanismu, který je však pro věrné zachycení reality nezbytný, externí validita tím také bude narušena.

Abstrakce je tak zřejmě tou vůbec nejnáročnější částí tvorby experimentálního designu. A to zejména proto, že to není úloha, kterou lze dělat podle návodu, ale dovednost, kterou je třeba nabýt praxí. Jakým způsobem k abstrakci budeme přistupovat, do velké míry záleží na výzkumných otázkách, které se pomocí experimentu snažíme zodpovědět. Metodologie sice určuje základní pravidla hry a určuje, co design obsahovat musí a co naopak obsahovat nesmí, avšak v rámci těchto pravidel existuje množství variant konečné podoby experimentu. Zde si můžeme pomoci porovnáním ke krasobruslení, ve kterém pravidla určují, jak dlouho má jízda trvat a jaké prvky musí obsahovat. Avšak to, jak bude výsledná jízda vypadat, záleží na zvolené hudbě a choreografii.

Pojďme se nyní přenést o několik dekád zpět do doby, kdy V. Smith a jeho kolegové (Smith et al., 1988, dále jen SSW) pracovali na vytvoření základního frameworku, který by jim umožnil zkoumat, zdali ceny, za které jsou aktiva na experimentálním trhu obchodována, odráží dostupné informace o jejich fundamentální hodnotě. Jinými slovy, zdali se účastníci na tomto experimentálním trhu, na kterém jim budou známy všechny potřebné informace, budou chovat racionálně. Tento cíl byl mimochodem zvolený velmi chytrě, protože se zaměřil na rozhodování v ideálních podmínkách, které v reálném světě nejsou nikdy dosaženy. Pokud by se v tomto ideálním světě začaly objevovat například tržní bubliny (a ony se objevovat začaly), bylo racionální předpokládat, že ve skutečném světě (ve kterém je mnohem více nejistoty a informace nejsou kompletní) bude tato dynamika ještě horší. Tento princip jakéhosi „minimálního benchmarku“ je v experimentálním designu často využíván (Gjerstad & Smith, 2014, s. 34).

SSW tedy stáli před úkolem zredukovat obchodování na burze do experimentálního designu, ve kterém účastníci mohou obchodovat s něčím, co jim po dobu několika experimentálních období (kol) bude přinášet zisk. Tento zisk musí být odhadnutelný, avšak jeho výše musí nabízet určitý rozptyl, který vytvoří prostor pro obchodování (kdyby byl zisk fixní a předem známý, pak by nemělo smysl aktivum obchodovat za cenu pod, nebo nad jeho fundamentální hodnotou). Základní framework tohoto experimentu pak vypadal následovně. Experiment měl 15 období (kol). Každá jednotka aktiva, kterou účastník v daném kole držel, náhodně vygenerovala

---

<sup>1</sup> To souvisí především s tím, že abychom byli schopni izolovat efekt každého měřeného aspektu rozhodování, musíme pro něj vytvořit vlastní experimentální skupinu. Jak se dočtete v kapitole 2.3, každá další skupina násobně zvyšuje nároky na počet účastníků, a tedy i na rozpočet studie.



jednu ze čtyř možných dividend – 0 centů, 8 centů, 28 centů nebo 60 centů. Protože pravděpodobnost získání každé z dividend byla stejná (tj.  $\frac{1}{4}$ ), očekávaná hodnota dividendy byla v každém kole 24 centů  $(0 + 8 + 28 + 60) / 4$ . Pokud participant tedy držel jednotku aktiva po celou dobu trvání experimentu, přinesla by mu dividendu v průměru  $0,24 \times 15 = \$3,60$ . Analogicky, fundamentální hodnota každé jednotky aktiva tedy byla počet zbývajících kol krát očekávaná hodnota dividendy (tj.  $\$3,60$  v prvním kole,  $\$3,36$  v druhém kole,  $\$3,12$  ve třetím kole, až  $\$0,24$  v posledním patnáctém kole). Každému participantovi byla na začátku přiřazena kombinace jedné, dvou, nebo tří jednotek aktiv a  $\$9,45$ ,  $\$5,85$ , nebo  $\$2,25$  hotovosti. Každý participant tak začínal s portfoliem v hodnotě  $\$13,05$ . V rámci každého časově omezeného kola pak participanti měli možnost mezi sebou obchodovat systémem dvojité aukce popsané výše (nyní však již anonymně pomocí počítače). Kdybych tedy byl například participant s portfoliem s vyšším zastoupením aktiv (tj. 3 jednotky aktiva a  $\$2,25$  hotovosti) a obával bych se, že dividendy budou v průměru nižší, než je jejich očekávaná hodnota, byl bych ochoten je v prvním kole prodat za cenu nižší, než je jejich fundamentální hodnota, tj.  $\$3,60$  za jednotku. Pokud by naopak jiný participant (s portfoliem s vyšším zastoupením hotovosti) věřil, že dividendy budou v průměru na úrovni očekávané hodnoty, nabídl by za jednotku aktiva právě  $\$3,60$ . Jiný participant by pak mohl dokonce věřit, že dividendy budou v průměru nad úrovní očekávané hodnoty a nabídl by tak za jednotku aktiva více než  $\$3,60$  (vzhledem k tomu, že již v této původní verzi experimentu vznikaly tržní bubliny, znamená to, že většina participantů byla tohoto typu).

Proces abstrakce je tedy především o nalezení její správné míry, protože jak vysoká, tak nízká přináší problémy spojené s externí validitou (tj. se zobecnitelností výsledků studie). Pokud nyní máte pocit, že experimentální ekonomie je komplikovaná a že snad ani nemá smysl to zkoušet, mějte na paměti, že to byl právě i tento experimentální design, za který Vernon Smith (který na designu pracoval od 50. let) obdržel Nobelovu cenu. Jak uvidíte v následujících kapitolách, mnoho designů je o poznání jednodušších. Drobné počáteční vystrašení je však na místě – je dobré si uvědomit, že vytvoření takto sofistikovaného designu není jen tak. Pokud jste začínající výzkumníci, je dobrou strategií využít nějakého existujícího a ověřeného frameworku (jakým je například tento **asset market**) a rozšířit ho o aspekt, který vás zajímá a nikdo ještě nezkoumal jeho vliv. Vaší kontrolní skupinou pak bude původní experiment od SSW, který v experimentální skupině obohatíte o to, co zajímá vás.

Kromě nevhodně zvolené míry abstrakce externí validitu ohrožují ještě další dva fenomény. Oba souvisejí s praktickými omezeními plynoucími ze skutečnosti, že je studie pořádána v experimentální laboratoři, a nikoliv ve **fieldu** (viz Boxík 13: Fieldový (terénní) experiment). Výhodou je, že – na rozdíl od abstrakce – na oba tyto problémy existují metodologické postupy, které více, či méně snižují negativní dopad na externí validitu. I z tohoto důvodu je každé problematice (stručně shrnuté v následujících dvou odstavcích) věnována samostatná kapitola.

Prvním problémem je, že laboratorní studie trvá řádově desítky minut a účastníci v rámci ní mohou disponovat pouze s prostředky, které jim nejdříve dáte. Ve skutečném světě však případné finanční dopady dobrého, nebo špatného rozhodnutí mohou mít mnohem zásadnější dopady, než že se vaše bohatství zvýší, nebo sníží o nižší stovky korun. Problematické mohou být také studie, ve kterých hraje roli čas (tj. výzkum časových preferencí). Může totiž být obtížné přimět stejné účastníky k účasti na studii skládající se z několika částí, které jsou odděleny delším časovým úsekem (například týdny nebo měsíce). Toto omezení je dobré mít na mysli již v počáteční fázi plánování výzkumu a zohlednit ho při rozhodování o tom, zdali je laboratorní experiment vhodnou metodou pro zodpovězení vašich výzkumných otázek.

Druhým problémem je skutečnost, že se experimentů účastní převážně studenti (často označováni akronymem WEIRD) univerzit, které disponují experimentálními laboratořemi. Tento problém bývá často označován za největší slabinu tohoto typu výzkumu. Někteří ho dokonce uvádějí jako důvod, proč by experimenty na studentských populacích neměly být vůbec realizovány, resp. proč by výsledky těchto studií neměly být publikovány. Tato kritika je oprávněná a na tento problém musí být brán zřetel především při interpretaci výsledků, avšak v žádném případě není důvodem pro zavržení laboratorního experimentálního výzkumu. Více o této problematice naleznete v kapitole 3.2.

**Boxík 13: Fieldový (terénní) experiment**

Fieldový experiment není realizován v laboratoři, ale v „terénu“. V kontextu behaviorální ekonomie je často označován zkratkou RCT, neboli Randomized Controlled Trial. V těchto kontrolovaných studiích jsou lidé náhodně a typicky bez svého vědomí zařazeni do jedné z výzkumných skupin. Příkladem je studie zaměřená na snížení předepisování antibiotik v Bostonu a Los Angeles (USA). Lékařům v experimentální skupině byl zaslán e-mail s informací, že předepisují více antibiotik, než je nezbytné. Výsledky ukázaly, že ve sledovaném období pak tito lékaři ve srovnání s kontrolní skupinou snížili předepisování antibiotik o 16 % (Meeker et al., 2016). Protože jsou fieldové experimenty prováděny na cílovém výzkumném vzorku, jejich externí validita má předpoklady být ve srovnání s laboratorními experimenty vyšší. Naopak kvůli omezené schopnosti výzkumníků kontrolovat experimentální prostředí může být snížena interní validita. Camerer (2011) však konstatuje, že při podobném nastavení studie jsou výsledky laboratorních experimentů srovnatelné s výsledky fieldových experimentů.

Typ experimentů	Relativní interní validita	Relativní externí validita	Limitace	Možnost replikace
Laboratorní	vysoká	nízká	výzkumy trvající delší dobu, velké částky, ztráty	vysoká
Fieldový	střední až vysoká	střední až vysoká	nemožnost zapojení/zásahu výzkumníka	nízká až střední

Tabulka 2: Porovnání experimentů v laboratoři a v terénu

Zdroj: Roe and Just (2009)

## 2.2 Interní validita

Zatímco externí validita je spojená spíše s designem experimentu, **interní validita** se týká převážně experimentální procedury a je definována jako *míra, s jakou jsou výzkumníci schopni prokázat kauzalitu pozorovaných vztahů* (Roe & Just, 2009) (viz Boxík 14: Korelace vs. kauzalita). Jinými slovy jde o to, zdali naměřené rozdíly mezi výzkumnými skupinami jsou způsobeny experimentální manipulací, nebo jinými okolnostmi studie. Jednou z takových okolností je například systematický rozdíl mezi výzkumnými skupinami zapříčiněný nevhodným rozdělením participantů do výzkumných skupin. Tomuto ohrožení interní validity se lze poměrně jednoduše vyhnout pomocí **randomizace**. Samotné rozhodování participantů pak může být ovlivněno faktory, kterých si výzkumníci nejsou vědomi. Efekt těchto faktorů pak mylně připsou experimentální manipulaci. Těmto faktorům se říká matoucí proměnné (confounding variables) a v laboratorních experimentech je jejich nejčastějším zdrojem **reaktivita**<sup>2</sup>.

### Boxík 14: Korelace vs. kauzalita

Korelace je vztah mezi dvěma proměnnými, které spolu mohou, ale nemusí souviset. Kauzalita nastává tehdy, když jedna proměnná přímo ovlivňuje druhou. Korelace však automaticky neznamená kauzalitu! Například mezi roky 1999 a 2009 lze v USA najít téměř perfektní korelaci (99,79 %) mezi počtem sebevražd oběšením a výdaji na vědu, vesmír a technologie. To však neznamená, že tento vztah je kauzální, tedy že Američané páchají více sebevražd v důsledku vyšších vládních výdajů.

(Vigen, 2022)

### 2.2.1 Randomizace

Randomizace je jedním ze základních konceptů celé experimentální metodologie a slouží ke snižování nevysvětlené variability v datech. Spočívá v zásadě, že kdykoliv v experimentálním designu nebo proceduře nemáme důvod pro specifickou organizaci, pak volíme organizaci náhodnou.

<sup>2</sup> Interní validita může být ohrožena ještě minimálně dvěma dalšími koncepty. Díky vysoké míře kontrolovatelnosti prostředí v laboratorních experimentech jsou však relevantní především ve fieldových experimentech. Prvním z těchto konceptů je „dozrávání“ (maturation) participantů. Pokud jsou participantů v delším časovém horizontu opakovaně vystavení stejné experimentální manipulaci, může se jejich reakce na ní měnit (například vlivem učení). Druhý koncept pak souvisí s neschopností výzkumníků spojit v čase experimentální manipulaci s reakcí participantů.

V experimentální proceduře se randomizace týká především rozřazování participantů do výzkumných skupin. Představte si, že chceme měřit vliv alkoholu na rozdíl v reakčních časech participantů. Participantům v experimentální skupině proto před zahájením testování dáme vypít deci 40% alkoholu, zatímco těm v kontrolní skupině dáme vypít deci vody. Jeden den pak sbíráme data do experimentální skupiny a druhý den sbíráme data do kontrolní skupiny. Ke svému překvapení však žádné rozdíly v reakčních časech nezjistíme. Důvodem však není neexistence efektu alkoholu na reakční čas, ale skutečnost, že se druhý den v kampusu konal pivní festival a většina participantů v kontrolní skupině se tak studie zúčastnila pod vlivem alkoholu. Kdybychom však participanty do výzkumných skupin přiřazovali v oba dny náhodně, pak by návštěvníci pivního festivalu byli rovnoměrně zastoupeni v kontrolní i experimentální skupině a my bychom tak efekt alkoholu na reakční čas objevili.

V experimentálním designu se randomizace týká především pořadí, ve kterém jsou participantům zobrazovány různé elementy studie. Mnohé výzkumy ukazují, že například pořadí otázek v dotazníku může ovlivnit odpovědi participantů. Pokud se lidí zeptáte na otázku „*Jak jste celkově spokojeni se svým životem?*“ a následně na otázku „*Na kolika rande jste byli minulý měsíc?*“, zřejmě mezi odpověďmi na obě otázky nenaleznete žádný vztah. Pokud se však na otázky zeptáte v opačném pořadí, odpovědi budou pozitivně korelovány – lidé, kteří byli na více rande, budou více spokojeni a lidé, kteří byli na méně rande, budou méně spokojeni (Strack et al., 1988). Podobně pokud byste se ptali na otázku „*Jakému nejzávažnějšímu problému nyní země čelí?*“ a poté na otázku „*Schvaluje, nebo neschvaluje způsob, jakým XY vykonává svou funkci premiéra?*“, zjistíte, že odpověď na druhou otázku bude do velké míry závislá na tom, jak premiér zvládá řešit problém, který respondent pokládá za nejzávažnější (Iyengar & Kinder, 1987). Pro experimentální metody pak jsou relevantnější situace, ve kterých participant odhadují nějaké hodnoty. Podle konceptu kotvení (anchoring) je možné, že pořadí, v jakém participant odhadují, může odhady ovlivnit (Epley & Gilovich, 2001). Například pokud se zeptáte na otázku „*Za kolik sekund uběhnete 100 metrů?*“, můžete dostat jiné odpovědi, než když se nejdříve zeptáte na otázku „*Jaký je světový rekord v běhu na 100 metrů?*“. Abyste se vyhnuli případnému efektu pořadí otázek, je dobré jejich pořadí randomizovat. Podle stejné logiky lze v případě multiple-choice otázek randomizovat pořadí jednotlivých možností. V případě within-subject designů pak lze randomizovat i pořadí výzkumných skupin (více se o tomto tématu dozvíte v kapitole 2.3).

### 2.2.2 Reaktivita

Reaktivita<sup>3</sup> zahrnuje několik dílčích konceptů, které obecně referují o změně rozhodování participantů v důsledku interakcí s experimentálním prostředím, výzkumníky a ostatními účastníky studie. Právě tyto interakce mohou být zdrojem různých matoucích proměnných, které mohou negativně ovlivnit výsledky studie. Cílem této části není poskytnout vyčerpávající přehled všech existujících typů reaktivity (jenž mnohdy mají v různých oborech jiné názvy), ale zmínit ty, které mají konkrétní implikace pro realizaci studie, resp. vedou ke konkrétním doporučením, jak jim předcházet.

#### A. Reaktivita jako důsledek experimentálních skupin

Prvním typem reaktivity je **Hawthornský efekt** (Hawthorne Effect), který upozorňuje na skutečnost, že participantí experimentů mohou spíše změnit své chování v důsledku vědomí, že je jejich chování pozorováno, než v důsledku experimentální manipulace (Wickström & Bendix, 2000). V laboratorních podmínkách je však tento efekt do velké míry neutralizován existencí kontrolní skupiny. Lze totiž předpokládat, že případné ovlivnění tímto efektem se projeví ve všech skupinách stejně a rozdíl mezi skupinami tak bude způsoben výhradně experimentální manipulací. Existence tohoto efektu je jedním z důvodů, proč je dobré mít ve studii vždy kontrolní skupinu (i přestože studie například navazuje na jiný výzkum, jehož výsledky by mohly posloužit jako kontrolní). Příbuzným konceptem, který je potenciálně větší hrozbou pro interní validitu, je **Efekt Johna Henryho** (John Henry effect). John Henry byl stavební dělník, který v druhé polovině 19. století za pomoci kladiva a ocelového vrtáku hloubil díry určené pro umístění výbušných náloží. Když se dozvěděl, že je jeho výkon porovnáván s parní vrtačkou, ve snaze dokázat svou efektivitu pracoval tak tvrdě, že v důsledku toho zemřel vyčerpáním (Saretsky, 1972). Efekt Johna Henryho tak v experimentální metodologii označuje situaci, ve které participantí mohou vyvíjet nadstandardní úsilí, pokud vědí, že jsou porovnávání s jinou skupinou a v čem se skupiny liší. Tento efekt je nutné brát v potaz při rozhodování, zdali design bude within-, nebo between-subjects (více v kapitole 2.3). Tento efekt může být také způsobený tím, že participantí, kteří se zúčastnili studie v kontrolní skupině, sdílí její obsah s participanty, kteří se studie teprve zúčastní a budou přiřazeni do experimentální skupiny. Tomu je velice těžké zabránit, proto je vždy lepší, když jsou data do studie nasbírána co možná nejrychleji a participantí tak nemají příliš času sdílet své zkušenosti. Dalším řešením je požádat participanty, aby se o obsahu studie až do jejího skončení s nikým nebavili.

<sup>3</sup> Koncept reaktivity pochází z psychologie a ve světě experimentální ekonomie je obsahem podobný koncept nazýván *Experimenter demand effects*. Protože tento název nemá překlad vhodný pro použití v tomto textu a zároveň nechceme zavádět nový termín, budeme se držet pojmu reaktivita.

## B. Reaktivita jako důsledek komunikace

Tento typ reaktivity vychází z nutnosti komunikace mezi experimentátorem a participanty, která může být implicitní (například ve formě řeči těla nebo vzezření), nebo explicitní (například ve formě verbálních, nebo psaných instrukcí). Příkladem ovlivnění studie implicitní komunikací je **efekt Chytrého Hanse**, koně, který na počátku 20. století v doprovodu svého majitele obrážel Německo a udivoval publikum svými neobvyklými dovednostmi. Hans dostal početní úlohu a kopytem označil správný výsledek, za což byl odměněn pamlskem. Následně provedená studie potvrdila, že se nejedná o podvod. Výzkumníci si ale všimli, že Hans odpoví správně pouze v případě, že zadavatel příkladu zná správný výsledek. Ukázalo se, že se Hans spíše než počítat naučil číst řeč těla – na základě postoje těla a výrazu v obličeji poznal, že na tabulce s výsledky narazil kopytem na ten správný (Ortmann, 2005). Pro ekonomické experimenty je pak relevantnější například studie (Ronay & Hippel, 2010), která prokázala, že přítomnost atraktivní ženy zvyšuje u mladých mužů hladinu testosteronu, v důsledku čehož jsou pak během studie ochotni více riskovat. Explicitní komunikace je v případě laboratorních experimentů spojena především s jejich organizací a udílením instrukcí. Protože výzkumníci typicky znají cíl studie, mohou participantům podvědomě poskytovat vodítka k tomu, jaké chování nebo rozhodování se v experimentu očekává.

Řešením tohoto problému je tzv. **zaslepení studie** (blinding), které má tři typy: jednoduché, dvojité a trojitě:

- **Jednoduše zaslepená studie** je taková, ve které participanti nevědí, zdali se účastní kontrolní, nebo experimentální skupiny. Ve většině laboratorních studií je jednoduché zaslepení běžným standardem.
- **Dvojitě zaslepená studie** je stejná jako jednoduše zaslepená studie s tím rozdílem, že ani výzkumníci realizující studii neví, do jaké skupiny je který participant zařazen. Pokud budeme pokračovat v paralele s testováním léků, pak lékař podávající experimentální látku subjektům neví, zdali se jedná o lék, nebo placebo. Tím je vyloučeno, že by subjekt mohl z implicitní, nebo explicitní komunikace poznat, do jaké skupiny je zařazen.
- **Trojitě zaslepená studie** je stejná jako dvojitě zaslepená studie s tím rozdílem, že výzkumníci analyzující experimentální data neví, jaká data pocházejí z kontrolní a jaká z experimentální skupiny. Tím je zabráněno souvisejícímu typu reaktivity spočívající v podvědomém ovlivnění statistické analýzy, například v důsledku sklonu k potvrzování (viz Boxík 15: Sklon k potvrzování (confirmation bias)). Vzhledem k tomu, že v dnešní době je již standardem tzv. předregistrace studií, trojitě zaslepená experimentální studie je zbytečnou komplikací

**Boxík 15: Sklon k potvrzování (confirmation bias)**

Ke sklonu k potvrzování dochází, když lidé vyhledávají a vyhodnocují informace tak, aby podpořili své stávající názory. Jedná se zřejmě o jeden z nejobustnějších biasů v teorii rozhodování, který lze najít v celé řadě odvětví od letectví (kde prokazatelně vedl k řadě tragických leteckých neštěstí), přes nákupní chování (kde lidé často vyhledávají pozitivní informace k produktům nebo značkám, které mají rádi a negativní informace k produktům nebo značkám, které rádi nemají), až po vědu (kde výzkumníci mohou nevědomě analyzovat a interpretovat data tak, aby našli podporu pro své hypotézy). Jednou z technik, kterou lze vliv sklonu k potvrzování zmenšit, je být sám sobě dáblovým advokátem – kromě přemýšlení o tom, proč je váš názor správný, se zamyslete také nad tím, proč by mohl být špatný.

(Oswald & Grosjean, 2004)

Připomeňme, že interní validita studie je ohrožena, pokud existují nesystémové rozdíly mezi skupinami, které mohou mimo jiné plynout z komunikace mezi experimentátory a účastníky. Je proto dobré, pokud je interakce omezena na nezbytné minimum. Toho lze dosáhnout tím, že bude co nejvíce instrukcí v **psané formě**. Je však zřejmé, že určitému množství interakce se nelze vyhnout – účastníky musí někdo přivítat v laboratoři, usadit je k počítačům a v průběhu studie být nablízku pro případ, že by nastal nějaký problém. Je proto dobré, pokud existuje tzv. **experimentální protokol**, neboli jakýsi scénář experimentu, který přesně definuje, jakým způsobem a jakými slovy se mají účastníci přivítat, jak má probíhat procedura usazování k počítačům (ta většinou zahrnuje randomizaci) a tak dále. Může také obsahovat doporučený dress-code pro případ, že sběr dat probíhá více dní (což většinou probíhá) a že se na administraci experimentu podílí více lidí (což většinou podílí).

Pro úplnost je třeba doplnit, že tento typ reaktivity je v literatuře označován jako *observer expectancy effect*, nicméně lze ho najít například i pod pojmy *observer effect* nebo *experimenter effect*. Z posledně jmenovaného je také odvozený český překlad – efekt experimentátora.



### C. Reaktivita jako důsledek očekávání participantů

Interní validita může být ohrožena i v případě, že experimentátor podnikne všechny kroky k tomu, aby se vyvaroval předchozím dvěma typům reaktivity. Další typ reaktivity je totiž způsoben očekáváním, které si participanti o studii vytvoří – nikoliv však v důsledku komunikace nebo drbů šířících se o studii, ale v důsledku snahy odhadnout, co se studie snaží zjistit. Participant se pak může stylizovat do role nápomocného participanta (good-participant), nebo do role zlomyslného participanta (negative-participant) (Weber & Cook, 1972). **Nápomocný participant** se snaží odhadnout záměr studie, aby svým chováním nebo rozhodováním přispěl k potvrzení hypotéz. Roli zde hraje skutečnost, že participanti nechtějí výzkumníkům studii „zkazit“. V literatuře lze tento koncept nalézt také pod označením *please-you effect*. **Zlomyslný participant** se naopak snaží svým chováním nebo rozhodováním přispět k vyvrácení hypotéz, aby studii zdiskreditoval. V literatuře lze tento koncept nalézt také pod označením *trhni-si efekt* (screw-you effect).

Tento typ reaktivity je jeden z důvodů, proč by design ekonomického experimentu měl obsahovat co nejméně **kontextu**. Tím se částečně vracíme k abstrakci – pokud je to možné, design experimentu by měl obsahovat pouze základní ekonomickou logiku zkoumaného mechanismu a oprostít se od kontextuálních detailů, které mohou participantům nabízet vodítka k odhadnutí účelu studie. Vrátime-li se k asset marketu od SSW, můžeme přemýšlet například o tom, že experimentální design neobsahuje slovo „akcie“, nýbrž „aktivum“. I přestože by se prakticky nic nezměnilo (akcie je přece druhem aktiva), mohlo by to u některých participantů vyvolat nechtěné asociace, jako například „obchodování s akciemi je nebezpečné, a proto to nebudu dělat“. Samozřejmě, že pokud půjdeme v této logice ještě o krok dál, můžeme namítnout, že i slovo aktivum může vyvolat nějaké asociace. Zde však již narážíme na to, že nelze použít žádné obecnější slovo, které zároveň zachová vnitřní logiku celého konceptu.

Protože již nějakou dobu přemýšlíte jako experimentální ekonom/ka, možná vás napadlo, že tento problém je do velké míry vyřešen tím, že existuje kontrolní skupina. V té by měl kontext, nebo jakákoliv jiná systematická vodítka hrát stejnou roli, a proto by neměly mít vliv na výsledky. To je pravda za podmínky, že jsou participanti **randomizováni**, neboli náhodně zařazeni do skupin (ideálně přímo experimentálním softwarem). Představte si například situaci, kdy by se nápomocní participanti hlásili na dřívější sessions (protože se již nemohou dočkat) a tyto sessions by byly pouze kontrolní skupina. Zlomyslní participant by se pak naopak hlásili na pozdější sessions (protože účast na experimentech je otrava, ale co by neudělali pro naštvání svého učitele-výzkumníka), které by byly pouze experimentální. V takovém případě by vznikl systematický rozdíl mezi skupinami a interní validita by byla narušena. Randomizace však zajistí, že obě skupiny budou obsahovat stejný počet nápomocných a zlomyslných participantů.

Vzhledem k existenci finančních incentív však lze očekávat, že převažující motivací většiny participantů ekonomických experimentů bude maximalizace odměny, kterou si za studii odnesou. I tak je třeba mít tento typ reaktivity v průběhu navrhování experimentu na mysli.

## D. Reaktivita jako důsledek společenských norem

Experiment polsko-amerického psychologa Solomona Asche poukázal na silnou potřebu lidí respektovat společenské normy (viz Boxík 16: Aschův test konformity). Poslední typ reaktivity souvisí právě s tímto fenoménem a týká se jak interakce participant-experimentátor, tak interakce participant-participant.

### Boxík 16: Aschův test konformity

Solomon Asch v 70. letech zkoumal, jak názory ostatních lidí ovlivňují rozhodování jednotlivců. Participantů tohoto experimentu měli za úkol nahlas říci, která ze tří svislých čar na jednom obrázku je stejně dlouhá jako čára na druhém obrázku. V každé experimentální session bylo vždy několik figurantů, kteří záměrně odpovídali špatně. Až jako poslední vždy odpovídal jediný skutečný participant. Ukázalo se, že 75 % participantů se podvolilo názoru většiny i přes to, že správná odpověď byla na první pohled patrná.

(Asch, 1951)

Pro ilustraci lze použít **Bradleyho efekt** (Bradley effect). Tento efekt popisuje situaci z kalifornských gubernérských voleb v roce 1982. Tom Bradley, starosta Los Angeles afro amerického původu, v těchto volbách kandidoval proti bělošskému kandidátovi Georgi Deukmejianovi. Bradley volby prohrál i přestože byl favorizován nejen průzkumy před volbami, ale i exit-polly (průzkum mezi voliči opouštějícími volební místnosti). Důvodem tak velkého zkreslení průzkumů byla zřejmě obava bělošských respondentů, že by mohli být nařčeni z rasismu (respektive, že by si to o nich tazatel mohl myslet) v případě, že za svého favorita označí Deukmejian. V anonymní volbě pak však zvolili právě kalifornského vrchního návladního (Payne, 2010). Tento efekt je pak projevem širšího konceptu, **sklonu ke společensky žádoucím chování** (social-desirability bias). Ten popisuje situaci, kdy se participant studie (nebo respondenti dotazníku) chovají v souladu se společenskými normami tak, aby se vyhnuli odsouzení ze strany druhých.

Řešením tohoto problému je, jak bylo zmíněno výše, **anonymita**. Pokud to povaha studie umožňuje, participant by během ní neměli být schopni identifikovat ostatní participanty, se kterými interagují. Výjimkou jsou samozřejmě studie, které se snaží vliv (ne)anonymity kvantifikovat. O poznání složitější je však zachování

naprosté anonymity vůči výzkumníkům. Participantům je zřejmé, že data jsou sbírána za účelem jejich analýzy, z čehož vyplývá, že k nim někdo má přístup. Z technických důvodů je pak vhodné mít data spárovaná s číslem počítače (místa) – to se u složitějších experimentů může hodit například v situaci, kdy během studie dojde k softwarovému problému, který je potřeba nějakým způsobem vyřešit. Z praktického hlediska je pak nutné přidělit participantovi jednoznačný identifikátor (což typicky bývá právě číslo počítače), podle kterého je na konci studie vyplácena odměna (ta je samozřejmě vyplácena oproti nacionalie). I přesto, že tento identifikátor není nikde propojen se jménem participanta a po jeho odchodu z laboratoře tyto dva údaje není možné zpětně spojit, participant může mít oprávněně<sup>4</sup> pocit, že to možné je. I přesto, že ve většině experimentálních studií (s výjimkou těch na společenské preference) nejsou sbírány žádné potenciálně citlivé údaje, je dobrou praxí participanty před začátkem studie v informovaném souhlasu upozornit, že sbíraná data jsou anonymní (v tom smyslu, že je zpětně nelze spojit s jejich osobou) a během studie se jich neptat na žádné osobní údaje (s výjimkou běžných demografických, jako je pohlaví či věk (viz Boxík 17: Pohlaví vs. gender)).

#### Boxík 17: Pohlaví vs. gender

Jedním z mála údajů, na které se během každého experimentu participantů ptáme, je jejich **pohlaví**. Pohlaví je chápáno výhradně v biologickém smyslu a na jeho základě se snažíme identifikovat rozdíly v chování a rozhodování, které mohou být způsobené odlišnými pohlavními orgány a s nimi spojenými predispozicemi. Například víme, že muži jsou ochotni riskovat více než ženy a že tento rozdíl je spojen s hladinou testosteronu. Proto participantům typicky dáváme na výběr z možností „mužské“ (nebo „muž“), „ženské“ (nebo „žena“) a „jiné“ (tato možnost zahrnuje především participanty s vrozenými defekty).

Pokud máme důvod domnívat se, že zkoumané chování by mohlo být spíše než biologii jednotlivců ovlivněno jejich vnímáním vlastní identity, pak je vhodnější ptát se na **gender**. V takovém případě byste však měli kromě možností „muž“ a „žena“ uvést také další relevantní možnosti, jako například „transgender“, „non-binary“ a další.

Protože v češtině může být „muž“ a „žena“ jak pohlaví, tak gender, dochází někdy k záměně těchto pojmů a jejich možností. Typicky tak, že v otázce na pohlaví jsou možnosti vztahující se k genderu. Ke správnému zapamatování rozdílů mezi těmito dvěma koncepty je užitečné podívat se na jejich anglické ekvivalenty. Zatímco pohlaví (sex) má v angličtině kromě „other“ možnosti „male“ a „female“ (české ekvivalenty samec a samice se u nás běžně nepoužívají ve spojitosti s lidmi), gender má v angličtině možnosti „man“, „woman“, „transgender“, „non-binary“ a tak dále.

<sup>4</sup> Výzkumníci by si teoreticky mohli během vyplácení čísla počítačů poznamenat do výplatního archu a data tak zpětně propojit s konkrétním participantem. Protože je však experimentální výzkum kvantitativní, neexistuje žádný důvod, proč by to měl dělat.

## 2.3 Multifaktoriální design

V této a další kapitole se budeme věnovat experimentálním designům, které mají dvě a více výzkumných skupin. Studie, ve které jsou do každé výzkumné skupiny přiřazení jiní participant, se označuje jako **between-subjects design** (zkráceně *between design*). Naopak studie, ve které se každý participant účastní více než jedné výzkumné skupiny, se označuje jako **within-subjects design** (zkráceně *within design*). Tyto názvy jsou odvozené od skutečnosti, zdali se výsledky jednotlivých skupin porovnávají „mezi“ výzkumnými vzorky nebo „v rámci“ stejného výzkumného vzorku. Jiný způsob, jak o těchto dvou možnostech přemýšlet, je, že zatímco v případě *between designu* je kontrolní skupina složena z jiných participantů než experimentální skupina (nebo skupiny), v případě *within designu* si je každý participant sám sobě kontrolní skupinou.

Ve Hře na důvěru (Trust Game) spolu hrají dva hráči, Investor a Správce. Oba na začátku hry obdrží prvotní vybavení ve výši  $e$  ECU. Investor posléze může zaslat libovolnou část svého prvotního vybavení Správci. Zasláná částka je však vynásobena předem určeným multiplifikátorem  $m$  a Správce tedy obdrží  $m$ -krát více, než kolik mu Investor zaslal. Správce poté může Investorovi zaslat zpět libovolnou část z obdržené částky (Berg et al., 1995). Například Investor i Správce na začátku hry obdrží 100 ECU. Část prvotního vybavení, kterou Investor zašle Správci, bude vynásobena multiplifikátorem  $m = 3$ . Investor se rozhodne Správci zaslat 50 ECU (který po vynásobení obdrží 150 ECU). Investorovi tedy zbyde 50 ECU a Správce má nyní k dispozici 250 ECU. Správce se nyní může rozhodnout, jakou část svého bohatství pošle Investorovi zpět. Typickým výsledkem Hry na důvěru je, že Investoři zasílají zhruba polovinu svého prvotního vybavení a Správci zasílají zhruba 37 % svého bohatství (Johnson & Mislin, 2011). V našem případě by tedy Správce zaslal Investorovi zpět zhruba 90 ECU (zaokrouhleno na desítky). Investor by tedy hru skončil se 140 ECU a Správce se 150 ECU.

Rozhodování Investorů a Správců však může záviset na parametrech hry, jakými jsou například výše multiplifikátoru, národnost participantů nebo výše prvotního vybavení. Zkusíme teď navrhnout experiment, ve kterém vliv těchto tří **faktorů** na rozhodování participantů ověříme. Faktor je nezávislá proměnná, kterou manipulujeme, abychom určili její vliv na měřenou, závislou proměnnou. Závislou proměnnou je v našem případě podíl prvotního vybavení zasláného Investorem a podíl bohatství zasláného Správcem zpět. Každý faktor pak má dvě a více **úrovní**. Úrovně jsou v jednoduchosti hodnoty, kterých faktor nabývá. Například náš faktor *Multiplifikátor* bude mít tři hodnoty:  $m = 3$  (jako v původní Hře na důvěru),  $m = 2$  a  $m = 4$ . Každá úroveň tak utvoří jednu výzkumnou skupinu. Protože nás zajímá, jak se rozhodování participantů změní v porovnání s původní verzí Hry na důvěru, skupinu s  $m = 3$  označíme jako **kontrolní** a zbylé dvě skupiny jako **experimentální**.

Výsledky obou experimentálních skupin pak porovnáme s výsledky kontrolní skupiny. Následující Tabulka 3 shrnuje faktor *Multiplikátor*, jeho úrovně a jejich označení (X, A, B).

	<i>kontrolní skupina</i>	<i>experimentální skupiny</i>	
	Úroveň m = 3	Úroveň m = 2	Úroveň m = 4
<b>faktor Multiplikátor</b>	X	A	B

Tabulka 3: Faktor *Multiplikátor* a jeho úrovně

Druhý faktor, který nás zajímá, je vliv národnosti participantů (*Národnost*). Kromě České republiky (CZE) budeme sbírat data také ve Vietnamu (VNM) a Ghaně (GHA). Protože nás u tohoto faktoru zajímá porovnání všech úrovní mezi sebou, nedává smysl výzkumné skupiny rozdělovat na kontrolní a experimentální. Následující Tabulka 4 shrnuje faktor *Národnost*, jeho úrovně a jejich označení (1, 2, 3).

	<i>výzkumné skupiny</i>		
	Úroveň CZE	Úroveň VNM	Úroveň GHA
<b>faktor Národnost</b>	1	2	3

Tabulka 4: Faktor *Národnost* a jeho úrovně

Abychom získali lepší přehled o celkovém počtu výzkumných skupin, sloučíme předchozí dvě tabulky do jedné (Tabulka 5). Tím vytvoříme tzv. 3x3 faktoriální design (v tomto zápise počet číslic reprezentuje počet faktorů a čísla samotná reprezentují počet úrovní každého faktoru). Vzniklá tabulka bude mít celkem  $3 \times 3 = 9$  buněk. Každá buňka pak reprezentuje jednu výzkumnou skupinu. Při pohledu na Tabulku 5 je důležité si uvědomit, že mezi sebou můžeme porovnávat výsledky pouze těch výzkumných skupin, které jsou označeny buď stejným písmenem (na každé úrovni *Multiplikátoru* můžeme porovnat výsledky z jiných zemí), nebo stejným číslem (na každé úrovni *Národnosti* můžeme porovnat výsledky z her s odlišnými *multiplikátory*). To vyplývá ze skutečnosti, že porovnávané výzkumné skupiny se od sebe mohou lišit pouze v jedné úrovni. Díky tomu jsme schopni určit, zda byl rozdíl (nebo jeho absence) mezi skupinami zapříčiněn *Národností*, nebo *Multiplikátorem*. Například má smysl porovnávat rozdíl v podílech prvotního vybavení zaslaných českými Investory, pokud se *multiplikátor* změní ze 3 na 2. Stejně tak můžeme porovnat rozdíl v podílech, které zasílají čeští a vietnamští Investoři, pokud je *multiplikátor* 3. V prvním případě pak můžeme rozdíl v zaslaných podílech jednoznačně přiřadit rozdílné výši *multiplikátoru* a v druhém případě můžeme rozdíl jednoznačně přisoudit odlišné národnosti investorů. Co naopak nemá smysl, je porovnávat rozdíl v zaslaných podílech mezi českými investory ve hře s *multiplikátorem* 2 a ghanskými investory ve hře s *multiplikátorem* 4.

Nevěděli bychom totiž, zdali je za zjištěný rozdíl zodpovědná národnost, nebo výše multiplikátoru.

	faktor Multiplikátor	Úroveň m = 3	Úroveň m = 2	Úroveň m = 4
faktor Národnost	Úroveň CZE	X1	A1	B1
	Úroveň VNM	X2	A2	B2
	Úroveň GHA	X3	A3	B3

Tabulka 5: 3x3 faktoriální design

Vlivu změny úrovně v rámci jednoho faktoru na závislou proměnnou říkáme **hlavní efekt** (main effect). V našem designu tak budeme mít dva hlavní efekty – vliv změny multiplikátoru na zaslaný podíl a vliv „změny“<sup>5</sup> národnosti na zaslaný podíl. Výhodou faktoriálních designů je, že nám kromě hlavních efektů umožňují zkoumat také **interakční efekty** (interaction effects). O těch hovoříme, pokud rozdíly, které způsobuje jeden faktor, závisí na úrovni druhého faktoru. Zatímco hlavní efekt je způsoben rozdíly mezi úrovněmi, interakční efekt je způsoben rozdíly mezi faktory. Například vliv změny multiplikátoru na zaslané podíly může být závislý na národnosti participantů. Lze si například představit, že výše multiplikátoru nemá žádný vliv v České republice, ve Vietnamu se s růstem multiplikátoru zaslané podíly snižují a v Ghaně se s růstem multiplikátoru zaslané podíly zvyšují.

V našem designu však ještě stále chybí jeden faktor, jehož vliv jsme chtěli prozkoumat. Tímto faktorem je výše prvotního vybavení a zajímá nás, zdali jeho zvýšení bude mít vliv na výši zaslaných podílů. Faktor *Prvotní vybavení* tedy bude mít dvě úrovně, a sice  $e = 100$  a  $e = 1000$ . Protože první úroveň je stejná jako ve všech předchozích výzkumných skupinách, ponecháme ji bez označení. Druhou úroveň budeme značit symbolem +. Jelikož tři faktory nelze zapsat do dvourozměrné tabulky, vyřešíme to tím, že každá buňka z předchozí tabulky bude obsahovat dvě výzkumné skupiny (tj. jednu skupinu pro každou ze dvou úrovní faktoru *Prvotní vybavení*). Tabulka 6 ilustruje vzniklý 3x3x2 faktoriální design s 18 výzkumnými skupinami.

	faktor Multiplikátor	Úroveň m = 3	Úroveň m = 2	Úroveň m = 4
faktor Národnost	Úroveň CZE	X1, X1+	A1, A1+	B1, B1+
	Úroveň VNM	X2, X2+	A2, A2+	B2, B2+
	Úroveň GHA	X3, X3+	A3, A3+	B3, B3+

Tabulka 6: 3x3x2 faktoriální design

<sup>5</sup> Národnost participantů se během experimentu samozřejmě nemění. O rozdílech způsobených tímto faktorem je však praktické přemýšlet jako o změně v zaslaném podílu, pokud české participanty zaměníme za vietnamské nebo ghanské.

Tímto způsobem lze samozřejmě přidávat další faktory, avšak s každým dalším faktorem násobně roste počet výzkumných skupin studie. Dále by nás mohlo například zajímat, jak se výsledky změní, pokud participanti budou znát pohlaví druhého hráče. Naši hypotézou by mohlo být, že participanti budou štedřejší k partnerům opačného pohlaví. Do studie bychom tak přidali faktor *Pohlaví* se třemi úrovněmi Neznámé, Stejně a Opačné. První úroveň by v tomto faktoru byla kontrolní skupina, ve které by participanti neznali pohlaví druhého hráče. Druhá a třetí úroveň by byly experimentální skupiny – v první je partner vždy stejného pohlaví a v druhé je partner vždy opačného pohlaví. Tím bychom vytvořili  $3 \times 3 \times 2 \times 3$  faktoriální design, který by celkem obsahoval 54 výzkumných skupin. Pokud bychom chtěli náš design zapsat do tabulky, při přidání čtvrtého faktoru již může být přehlednější z každé jeho úrovně udělat blok, který bude mít vlastní (stejnou) tabulku. Tabulku 6 bychom tak zvlášť měli pro úroveň Neznámé, pro úroveň Stejně a konečně také pro úroveň Opačné. Lze však říci, zapisování více než 3-faktorového designu do tabulek již nepřináší velkou přidanou hodnotu.

## 2.4 Smíšený design

Pokud bychom chtěli opravdu realizovat výzkum z Tabulky 5, museli bychom každou z 9 skupin naplnit participanty. O tom, jak určit, kolik participantů by v každé skupině mělo být, se dočtete v kapitole 3.3. Pro ilustraci nyní postačí, když budeme pro každou výzkumnou skupinu počítat s 30 participanty, z nichž polovině bude náhodně přidělena role Investora a polovině role Správce. Na realizaci celé studie bychom tak potřebovali dohromady  $9 \times 30 = 270$  participantů. Při show-up fee ve výši 150 Kč a prvotním vybavení ve výši 100 Kč<sup>6</sup> by náklady studie byly zhruba 81 000 Kč<sup>7</sup>. Abychom minimalizovali počet participantů nutný pro realizaci studie (a tím i její rozpočet), můžeme přemýšlet o transformaci between designu na within. Pokud, tak jako v našem příkladě, některý z faktorů transformovat nelze, studie bude mít **smíšený design**. Ten obsahuje jeden nebo více between design faktorů a jeden nebo více within design faktorů. Je zřejmé, že faktor *Národnost* nemůže být transformován do within verze, protože není možné, aby jeden participant měl během experimentu všechny tři národnosti. Oproti tomu faktor *Multiplikátor* do within verze být transformován může, protože jeden participant může hrát hru třikrát po sobě, po každé s jiným multiplikátorem. V takovém případě hovoříme o tom, že se participant v rámci jednoho experimentu zúčastní více kol (v tomto

<sup>6</sup> Vzhledem k tomu, že ve Hře na důvěru je zasláná část prvotního vybavení vynásobena multiplikátorem, není problém, pokud bude prvotní vybavení nižší než show-up fee.

<sup>7</sup> Každý z 270 participantů obdrží show-up fee ve výši 150 Kč. Každý Správce obdrží prvotní vybavení ve výši 100 Kč, nicméně na základě výsledků předchozích studií očekáváme, že polovinu odešle Správci a ta tak bude vynásobena průměrným multiplikátorem  $(2+3+4)/3 = 3$ . To znamená, že  $270 \times 150 + 135 \times 100 + 135 \times (0,5 \times 100 + 0,5 \times 100 \times 3) = 81\,000$ .

případě tří)<sup>8</sup>. Z pohledu do upravené Tabulky 7 je patrné, že díky transformaci z between designu do smíšeného designu jsme z původně 9 výzkumných skupin udělali pouze 3. Pokud by nám tedy stále stačilo mít v každé skupině 30 participantů, pak jsme celkový počet participantů snížili z 270 na 90 a rozpočet z 81 000 Kč na 27 000 Kč.

	faktor Multiplikátor	kola		
		Úroveň m = 3	Úroveň m = 2	Úroveň m = 4
faktor Národnost	Úroveň CZE	1		
	Úroveň VNM	2		
	Úroveň GHA	3		

Tabulka 7: Smíšený design

Z pohledu rozpočtu je tedy within design vždy preferovanější variantou než between design. Umožňuje totiž, *ceteris paribus*, nasbírat stejné množství dat za méně peněz. Proč tedy nejsou všechny studie within design? V některých situacích to zkrátka není možné – příkladem je faktor *Národnost*. V jiných situacích pak within design může ohrožovat interní validitu studie. To platí především pro studie, ve kterých je zřejmý rozdíl mezi kontrolní a experimentální skupinou. Participantů pak mohou snadněji odhadnout, jaký cíl experiment sleduje a upravit podle toho své rozhodování. To by mohl být příklad výše zmíněného faktoru *Pohlaví*. Pokud bychom tento faktor realizovali jako within design, participantů by mohli dbát na to, aby zasílali stejné částky bez ohledu na to, zdali je druhý hráč neznámého, stejného, nebo opačného pohlaví. Důvodem pro jejich chování by však nemusely být jejich preference, ale reaktivita v důsledku společenských norem (v tomto případě spojených s rovností pohlaví). V kontextu jiné studie by mohly být výsledky ovlivněny i jinými typy reaktivity. Pro within design jsou tedy vhodné ty faktory, u kterých si můžeme být relativně jistí, že znalost všech jejich úrovní neovlivní rozhodování participantů. Z tohoto pohledu je dobrým příkladem faktor *Multiplikátor*. U toho lze očekávat, že s vyšší multiplikátoru ve Hře na důvěru nejsou spojeny žádné společenské, ani jiné normy, které by vedly k reaktivitě systematicky ovlivňující celkové výsledky<sup>9</sup>. V rámci jakési předběžné opatrnosti pak lze začínajícím výzkumníkům doporučit, aby využití within designu důkladně konzultovali se zkušenějšími kolegy.

<sup>8</sup> Otázkou může být, v jakém pořadí mají jednotliví participantů kola hrát. Univerzální radou je, aby pořadí kol bylo randomizováno, tedy aby každý participant hrál kola v náhodném pořadí. Mohou ovšem nastat okolnosti, za kterých například můžete chtít, aby první kolo bylo vždy kontrolní skupina a randomizovat až pořadí experimentálních skupin.

<sup>9</sup> Lze si například představit, že několik participantů se znalostí předchozích výsledků Hry na důvěru s odlišnými multiplikátory se bude snažit tyto výsledky svými odpověďmi „replikovat“. Pokud se však náš vzorek neskládá pouze ze studentů předmětu Experimentální ekonomie, kteří se o této hře zrovna minulý týden učili, pak bude vliv tohoto typu reaktivity zanedbatelný.



Kromě nižších nároků na rozpočet within design typicky generuje kvalitnější data (pokud se nám podaří vyhnout nástrahám spojených s interní validitou). Rozdíly v rozhodování mezi jednotlivými výzkumnými skupinami jsou totiž v rámci rozhodování stejného participanta způsobeny spíše experimentální manipulací než náhodným šumem. **Náhodný šum** (random noise) je, podle definice, nevysvětlitelná variabilita v datech. Lépe tento koncept však pochopíte na následujícím příkladu. Představte si, že rozhodnutí Investorů o tom, jaký podíl ze svého prvotního vybavení zašlou Správci, závisí kromě výše multiplikátoru (2, 3, nebo 4) i na jejich náladě. Pokud mají Investoři špatnou náladu, zasílají méně, než pokud mají dobrou náladu. I přes správnou randomizaci participantů do výzkumných skupin by se nám mohlo přihodit (obzvláště pokud náš výzkumný vzorek není velký), že ve skupině s  $m = 4$  budeme mít více participantů se špatnou náladou než ve zbylých dvou skupinách. Ve výsledku tak zjistíme, že zatímco změna multiplikátoru z 2 na 3 vede ke zvýšení zaslaného podílu, změna ze 3 na 4 ke zvýšení nevede. Protože však efekt nálady na zasílané podíly nekontrolujeme (nedali jsme participantům žádný dotazník indikující jejich náladu), můžeme se mylně domnívat, že tento výsledek byl způsoben výší multiplikátoru. Právě nálada je tedy v tomto příkladu zdrojem náhodného šumu. Obecně jde říci, že náhodný šum způsobuje cokoliv, co ovlivňuje rozhodování participantů, ale není naším designem kontrolováno. Pokud se však participant zúčastní všech tří výzkumných skupin, vliv nálady na variabilitu dat v rámci jejich odpovědí bude neutralizován (za předpokladu, že se nálada participantů nemění mezi jednotlivými koly). Protože chceme zjistit, jak změna multiplikátoru ovlivní změnu zaslaného podílu, nemusíme si dělat starosti s tím, že nálada může ovlivňovat rozdíly v nominálních částkách, které jednotliví Investoři zasílají Správcům. Například Investor s dobrou náladou zašle 20 % svého prvotního vybavení při  $m = 2$ , 40 % při  $m = 3$  a 80 % při  $m = 4$ . Investor se špatnou náladou zašle 10 % svého prvotního vybavení při  $m = 2$ , 20 % při  $m = 3$  a 40 % při  $m = 4$ . U obou Investorů zvýšení multiplikátoru o jednotku vede ke zdvojnásobení zaslaného podílu, i přestože ti se špatnou náladou zasílají Správcům podíly na nominálně nižší úrovni.

Tento příklad je shrnutý v následujících dvou tabulkách. První Tabulka 8 znázorňuje situaci, kdy je výzkum realizován jako between design. V každé ze tří skupin jsme nasbírali data od 10 různých participantů. První řádek tabulky ukazuje, jaký podíl prvotního vybavení každý z Investorů zaslal svému Správci. Všimněte si, že zatímco ve skupině s multiplikátory 2 a 3 mají Investoři dobrou náladu, ve skupině s multiplikátorem 4 mají Investoři špatnou náladu. Proto se průměr druhé a třetí skupiny neliší a je shodně 40 %. Průměr první skupiny je 20 %. Poslední řádek tabulky ukazuje rozdíly mezi skupinami. Zatímco rozdíl mezi první a druhou skupinou je 100 %, rozdíl mezi druhou a třetí skupinou je 0 %.

	<b>m = 2</b> <b>Investoři s</b> <b>dobrou náladou</b>	<b>m = 3</b> <b>Investoři s</b> <b>dobrou náladou</b>	<b>m = 4</b> <b>Investoři se</b> <b>špatnou náladou</b>
<b>Podíly prvotního vybavení (v %)</b>	10, 0, 40, 30, 20, 20, 30, 10, 10, 10	40, 30, 60, 40, 30, 50, 40, 40, 40, 30	60, 40, 10, 50, 40, 40, 40, 40, 50, 30
<b>Průměr</b>	20 %	40 %	40 %
<b>Rozdíl (<math>\Delta</math>)</b>	100 %		0 %

Tabulka 8: Between design

Druhá Tabulka 9 znázorňuje situaci, kdy je výzkum realizován jako within design. Pokud by se výzkumu zúčastnili stejní participantů se stejnou náladou, pak bychom nasbírali odpovědi od 20 Investorů s dobrou náladou a od 10 Investorů se špatnou náladou. Investoři s dobrou náladou stejně jako v předchozí tabulce zasílají 20 % a 40 % svého prvotního vybavení při multiplikátorech 2 a 3. Pokud je multiplikátor 4, investoři s dobrou náladou přispívají 80 % a rozdíl mezi jednotlivými multiplikátory je vždy 100 %. Investoři se špatnou náladou stejně jako v předchozí tabulce zasílají 40 % při multiplikátoru 4. Protože mají špatnou náladu, zasílají méně i při multiplikátorech 2 a 3, konkrétně 10 % a 20 %. Jak však ukazuje poslední řádek, rozdíl mezi jednotlivými multiplikátory je stejně jako u Investorů s dobrou náladou také 100 %.

	<b>m = 2</b>	<b>m = 3</b>	<b>m = 4</b>
<b>Investoři s dobrou náladou</b>	10, 0, 40, 30, 20, 20, 30, 10, 10, 10, 10, 20, 10, 20, 20, 10, 20, 10	40, 30, 60, 40, 30, 50, 40, 40, 40, 30, 30, 30, 30, 40, 40, 50, 50, 40, 60, 60	80, 90, 80, 70, 60, 100, 90, 90, 80, 80, 80, 80, 90, 60, 80, 70, 70, 100, 100, 70
<b>Průměr</b>	20 %	40 %	80 %
<b>Rozdíl (<math>\Delta</math>)</b>	100 %		100 %
<b>Investoři se špatnou náladou</b>	10, 10, 20, 20, 20, 0, 10, 10, 0, 10	20, 0, 20, 0, 30, 40, 20, 40, 10, 20	60, 40, 10, 50, 40, 40, 40, 40, 50, 30
<b>Průměr</b>	10 %	20 %	40 %
<b>Rozdíl (<math>\Delta</math>)</b>	100 %		100 %

Tabulka 9: Within design

Čenou za kvalitnější data u within designů je také o něco náročnější statistická analýza, která musí brát v potaz skutečnost, že jeden participant poskytuje více odpovědí.

Následující Tabulka 10 shrnuje relativní výhody a nevýhody between a within designu. Between design méně ohrožuje interní validitu studie a klade nižší nároky na statistickou analýzu dat. Jeho nevýhodou však je nutnost účasti většího počtu participantů (což se negativně projevuje na výši rozpočtu a organizační náročnosti studie), jejichž data navíc poskytují méně přesné výsledky v důsledku náhodného šumu. Within design naopak poskytuje přesnější výsledky s nižšími nároky na počet participantů (tedy i na rozpočet a organizaci). Velkou nevýhodou, která potenciálně může kompromitovat výsledky celé studie, je možné ohrožení interní validity v důsledku nezachycené reaktivity. Drobnou nevýhodou pak jsou také vyšší nároky na analýzu dat.

Faktory	Between design	Within design
nároky na počet participantů/ rozpočet	vyšší (-)	nižší (+)
ohrožení interní validity	nižší (+)	vyšší (-)
nároky na analýzu dat	nižší (+)	vyšší (-)
přesnost výsledků	nižší (-)	vyšší (+)

Tabulka 10: Porovnání between a within designu

## 2.5 Vícekolové experimenty

Hra na důvěru je typickým příkladem experimentu, který je realizován jako vícekolový. Pokud by se totiž hrálo jenom jedno kolo (one-shot game), Správci by neměli žádný důvod zasílat Investorům zpět jakoukoliv částku vyšší než 0. Protože si toho jsou Investoři dobře vědomi, Správcům nezašlou nic. Hra tak skončí ve svém Nashově ekvilibriu (viz Boxík 18: Nashova rovnováha). Ve chvíli, kdy však má hra více kol, Správci i Investoři se začnou chovat strategicky – obě strany totiž ví, že ze vzájemné spolupráce mohou benefitovat. Nezáleží přitom ani na tom, zdali má hra ve všech kolech stejné parametry, nebo se některé parametry (například multiplikátor) mezi koly mění. Důležité je, aby všechna kola hrála stejná – byť anonymní – dvojice participantů, mezi kterými tak může vzniknout důvěra. Protože participantů většinou vědí, kolik kol hra bude mít, s blížícím se koncem experimentu se spolupráce začne zhoršovat a poslední kolo je pak často hráno, jako kdyby byla hra jednokolová<sup>10</sup>. Kromě studií, které obsahují element strategického rozhodování, se vícekolové experimenty používají například ve studiích vyžadujících učení nebo zkoumajících roli zkušeností. Jak také vyplývá z předchozí kapitoly, jakýkoliv within design experiment je ze své podstaty vícekolový.

<sup>10</sup> Je sice možné participantům například říci, že počet kol bude určen náhodně a bude se pohybovat mezi X a Y koly, nicméně lze očekávat, že tato nejistota může negativně ovlivnit jejich spolupráci ve všech kolech, která mohou být poslední.

**Boxík 18: Nashova rovnováha**

Nashova rovnováha (Nash equilibrium) nastává tehdy, pokud žádný z hráčů není motivován změnit svou strategii při uvážení strategií všech ostatních hráčů (Holt & Roth, 2004). Představte si například křižovatku, na kterou se z obou stran blíží auto. Každý řidič může buď zastavit, nebo jet dál. Pokud oba řidiči zvolí stejnou strategii (oba zastaví, nebo oba pojedou dál), budou na tom o poznání hůře, než pokud zvolí opačnou strategii (první zastaví a druhý pojedou dál, nebo naopak). Tato „hra“ má tedy dvě Nashovy rovnováhy. Semaforey pak pomáhají náhodně a spravedlivě určit, který řidič má zastavit a který má jet dál. Motivace řidičů k respektování světelných signálů pak ani tak nepramení ze strachu z pokuty, ale z vědomí, že změna strategie (jet na červenou, nebo stát na zelenou) nepovede k lepšímu výsledku.

Nashova rovnováha je pojmenována po americkém matematikovi Johnu Nashovi, který v roce 1994 obdržel Nobelovu cenu za ekonomii za svou práci v oblasti teorie her (zajímavostí je, že cenu obdržel společně s jedním z evropských průkopníků experimentální ekonomie, Reinhardem Seltenem). O jeho životě, který byl poznamenán schizofrenií a její léčbou, byl natočen film Čistá duše (A Beautiful Mind). Nash zemřel v roce 2015 při autonehodě ve věku 86 let

Až na výjimky je u vícekolových experimentů záhodno, aby rozhodování v každém kole bylo co možná nejvíce nezávislé. Tím je v kontextu experimentální ekonomie myšleno především to, aby výše variabilní odměny za dané kolo závisela pouze na rozhodování v tomto kole. Prvotní vybavení tak musí být na začátku každého kola pro všechny participanty stejné a nezávislé na jejich rozhodování v předcházejících kolech. Na konci experimentu je pak náhodně<sup>11</sup> vybráno jedno kolo, za které je participantům vyplacena odměna. Kdyby participanté obdrželi odměnu za každé kolo experimentu<sup>12</sup>, mohl by jejich rozhodování ovlivnit **efekt bohatství** (wealth effect). Například participanté se mohou účastnit experimentu s konkrétní představou o tom, kolik peněz si účastí chtějí vydělat (popřípadě si tuto představu mohou vytvořit během čtení instrukcí). Pokud by pak v průběhu experimentu této částky dosáhli, mohli by se s tím spokojit a v dalších kolech se chovat zvláštně – například zasílat vysoké částky Správcům i přestože ti jim přestali zasílat cokoliv zpět. Takovéto chování Správců by pak mohlo být vysvětleno tím, že ti naopak stanovené částky ještě nedosáhli.

<sup>11</sup> Participanté často nevěří, že náhodně vybrané kolo bylo vybráno zcela náhodně – obzvlášť pokud vybrané kolo je zrovna to, ve kterém je jejich odměna nejvyšší. Je proto dobré, aby byl proces výběru náhodného čísla spojen s nějakou aktivitou samotného participanta. Můžete ho například nechat „hodit“ vícestrannou digitální kostkou nebo mu pusit sekvenci rychle měnících se čísel 1–10 a mezerníkem ho nechat sekvenci zastavit.

<sup>12</sup> V takovém případě byste navíc museli snížit konverzní poměr mezi ECU a lokální měnou.

Doposud jsme se také zabývali pouze situacemi, kdy každý participant v rámci jednoho činí pouze jedno rozhodnutí. Pokud je však experiment komplexnější, může obsahovat více než jedno rozhodnutí. Na začátku každého kola Hry na důvěru bychom například mohli nechat participanty odhadnout velikost podílu, který jim druhý hráč zašle. Tento odhad by samozřejmě musel také být incentivizovaný<sup>13</sup>. Pokud je odměna za odhad relativně vysoká vzhledem k prvotnímu vybavení, pak hrozí, že se participanti budou snažit zasílat takové částky, které maximalizují jejich šance na získání odměny za odhad. Například Investoři vědí, že pokud nezašlou Správcům nic, pak jistě obdrží odměnu za svůj odhad, že jim Správci také zašlou 0 %. Jedním řešením tohoto problému je, že odměny za odhad budou relativně nízké a participantům se tedy nevyplatí cílit své odhady. Druhým řešením pak je náhodné vyplácení buď odměny za odhad, nebo odměny za samotnou Hru na důvěru.

Za svým způsobem vícekolové mohou být považovány také experimenty, které mají jen jedno kolo, které však obsahuje větší množství rozhodnutí. Typickým příkladem jsou studie zaměřené na elicitaci rizikových preferencí (například Holt and Laury (2002)), ve kterých participanti dostanou několik párových loterií a jejich úkolem je si z každého páru vybrat možnost A, nebo možnost B. Každý pár vypadá zhruba takto. Možnost A přináší 10% šanci získat 100 ECU a 90% šanci získat 80 ECU. Možnost B přináší 10% šanci získat 190 ECU a 90% šanci získat 5 ECU. Po zvolení jedné z možností se participanti přesunou k dalšímu páru možností, který se liší svou rizikovostí a vyšší odměny. Pokud by participanti dostali odměnu za každý pár loterií, pak by jejich rozhodování mohlo být ovlivněno **portfolio efektem**. Zatímco v některých párech by volili rizikovější možnosti, v jiných by mohli volit méně rizikové možnosti. Touto diverzifikací svého „portfolia“ by se snažili vyhnout získání nízké konečné odměny. I zde je tak řešením náhodné vyplácení pouze jednoho páru loterií.

<sup>13</sup> V nejjednodušší variantě lze použít fixní odměnu za odhad, který se třetí do stanoveného tolerančního pásma. Sofistikovanější mechanismus je například Quadratic Scoring Rule, který vyplácí tím větší odměnu, čím přesnější odhad je. Více informací o tomto mechanismu najdete například v článku (Sonnemans & Offerman, 2001).

## 2.6 Shrnutí

### Externí validita a abstrakce

Externí validita je míra, s jakou lze výsledky laboratorního výzkumu použít ve skutečném světě. Může být narušena především nevhodnou mírou abstrakce při vytváření experimentálního designu, neodpovídající výši incentív, časovými prodlevami mezi jednotlivými částmi studie nebo složením výzkumného vzorku.

### Interní validita, randomizace a reaktivita

Interní validita je míra, s jakou jsou výzkumníci schopni prokázat kauzalitu pozorovaných vztahů. Stejně jako u externí validity i interní validita by měla být co možná nejvyšší. Interní validita může být ohrožena nedostatečnou randomizací, reaktivitou, učněním nebo nejasnými instrukcemi. **Randomizace** slouží ke snižování nevysvětlené variability v datech a spočívá v zásadě, že kdykoliv v experimentálním designu nebo proceduře nemáme důvod pro specifickou organizaci, pak volíme organizaci náhodnou. V experimentální proceduře se randomizace týká především rozřazování participantů do výzkumných skupin. V experimentálním designu se randomizace týká především pořadí, ve kterém jsou participantům zobrazovány různé elementy studie. **Reaktivita** zahrnuje několik dílčích konceptů, které obecně referují o změně chování participantů v důsledku účasti v laboratorní experimentální studii. *Hawthornský efekt* a *Efekt Johna Henryho* popisuje změnu chování participantů v důsledku vědomí, že je jejich chování pozorováno. Řešením je existence kontrolní skupiny. *Efekt Chytrého Hanse* upozorňuje na vliv implicitní, či explicitní komunikace mezi výzkumníky a participanty při organizaci experimentu a udílení instrukcí. Řešením je zaslepení studie (jednoduché, dvojité nebo trojitě), psané instrukce a experimentální protokol. Dalším typem reaktivity je takzvaný *nápomocný (zlomyslný) participant*. Nápomocný (zlomyslný) participant se snaží rozhodovat tak, aby výzkumníkům poskytl (neposkytl) podporu pro domnělé hypotézy. Design experimentu proto má obsahovat co nejméně kontextu. *Bradleyho efekt* pak hrozí, pokud participanti upravují své chování tak, aby bylo v souladu se společenskými normami a oni se tak vyhnuli hrozícímu odsouzení ze strany druhých. Řešením je anonymita studie.

## Kontrolní a experimentální skupiny

**Faktor** je nezávislá proměnná, kterou manipulujeme, abychom určili její vliv na měřenou, závislou proměnnou. Každý faktor pak má dvě a více úrovní. **Úrovně** jsou hodnoty, kterých faktor nabývá. Každá úroveň tak tvoří jednu **výzkumnou skupinu**. **Kontrolní skupina** obsahuje základní design studie a její výsledky se porovnávají s výsledky **experimentální skupiny**, která je rozšířená o jeden sledovaný faktor. Experimentální design je možné rozšiřovat jak horizontálně, tak vertikálně. V rámci faktoriálního designu je možné zkoumat **hlavní a interakční efekt**. Zatímco hlavní efekt je způsoben rozdíly mezi úrovněmi, interakční efekt je způsoben rozdíly mezi faktory.

## Between-subjects design vs. Within-subjects design

Pokud se každé výzkumné skupiny účastní jiní participanti, hovoříme o **between-subjects designu**. Studie, ve které se jeden participant účastní dvou nebo více výzkumných skupin, se označují jako **within-subjects design**. Jedna studie může obsahovat jak between design, tak within design. V takovém případě pak hovoříme o **smíšeném designu**. Between design studie mívají vyšší interní validitu a nižší nároky na statistickou analýzu dat. Within design klade nižší nároky na počet participantů (a tedy i na větší rozpočtu) a poskytuje přesnější výsledky.

## Vícekolové experimenty

Vícekolové experimenty se používají například ve studiích vyžadujících učení, obsahujících element strategického rozhodování nebo zkoumajících roli zkušeností. Jakýkoliv within design experiment je ze své podstaty vícekolový. Výše variabilní odměny za dané kolo by měla vždy záviset pouze na rozhodování v daném kole. Na konci experimentu je pak náhodně vybráno jedno kolo, za které je participantům vyplacena odměna. V opačném případě by mohl jejich rozhodování ovlivnit efekt bohatství. Vícekolové mohou být také jednokolové experimenty obsahující větší množství rozhodnutí.





Experimentálních studií se typicky účastní vyšší desítky až nižší stovky participantů. V této souvislosti je třeba zodpovědět minimálně tři důležité otázky. První otázkou je, kde jako výzkumníci seženete takové množství participantů. Druhou otázkou je, jaké bude složení vašeho výzkumného vzorku, tedy jací participanti se výzkumu zúčastní. Třetí otázkou je, kolik participantů přesně potřebujete pro realizaci vašeho experimentu. Odpovědi na tyto otázky – které spolu často souvisí – nabízí tato kapitola.

### 3.1 Rekrutace participantů

Z mnoha praktických i teoretických důvodů chceme data nasbírat v co možná nejkratším časovém úseku. Délka sběru dat se odvíjí od kapacity experimentální laboratoře a délky studie. Protože ekonomické experimenty typicky netrvají déle než hodinu a jak bylo uvedeno výše, kapacita většiny laboratoří se pohybuje mezi 15 a 25 místy, lze za den teoreticky nasbírat data od zhruba 100 participantů. Prakticky však toto číslo bývá mnohem nižší, protože bývá obtížné naplnit všechny vypsané termíny.

Rekrutaci participantů většinou zajišťuje laboratoř, ve které experiment pořádáte. Na základě vašich požadavků a volných kapacit vypíše termíny, které rozešle lidem zaregistrovaným pro účast v experimentálních studiích. Tito lidé pak mají možnost přihlásit se na jeden z vypsanych termínů. Počet přihlášených pak závisí na mnoha faktorech. Mezi nejdůležitější patří **datum a čas termínu** (protože laboratoře jsou typicky v univerzitních kampuzech a většina participantů jsou studenti dané univerzity, je dobré, aby se studie konala během semestru a v časech, v kterých probíhá výuka) a **výše show-up fee** (které je uvedeno v pozvánce<sup>1</sup>). V závislosti na těchto dvou faktorech se k účasti ve studii přihlásí typicky 10–20 % pozvaných. Počet pozvaných je pak samozřejmě omezen celkovým počtem lidí zaregistrovaných v databázi.

Protože datum a čas termínu vždy někomu vyhovovat nebude a celkový počet lidí v databázi laboratoře svojí aktivitou moc nenavýšíte, jediným způsobem, kterým můžete ovlivnit počet přihlášených participantů, je výše show-up fee. Zde platí základní ekonomické pravidlo o nabídce a poptávce – čím méně peněz participantům

<sup>1</sup> V pozvánce lze uvádět i například průměrnou výši variabilní odměny. Avšak pozor! Touto informací můžete v participantech vytvořit očekávání, která mohou ovlivnit jejich rozhodování v průběhu studie. Kompromisním řešením tak je uvést výši show-up fee doplněnou o informaci, že v závislosti na svém rozhodování mohou participantů získat dodatečnou odměnu.

za účast nabídnete, tím menší část z nich se bude chtít studie zúčastnit<sup>2</sup>. S určením přesné výše show-up fee by vám měla pomoci laboratoř, která je na základě realizovaných výzkumů schopna určit vztah mezi výší show-up fee a počtem lidí, kteří se k účasti přihlásí. Protože již víme, že show-up fee by nemělo být vyšší než variabilní část odměny, každé jeho navýšení se do nákladů na jednoho participanta propíše hned dvakrát. V zásadě tak řešíte dilema mezi malou účastí v případě příliš nízkého show-up fee a nároky na rozpočet v případě příliš vysokého show-up fee.

### 3.2 Složení výzkumného vzorku

I přestože se do účastnických databází typicky může zaregistrovat kdokoliv starší 18 let, většina registrovaných jsou současní, či bývalí studenti univerzity. To je způsobeno několika důvody (Feltovich, 2011). Studenti jsou časově flexibilní a mohou se tak účastnit termínů studie, kterých by se běžně pracující lidé účastnit nemohli. Dalším důvodem je, že finanční odměna za účast bývá pro studenty atraktivnější než pro normálně vydělávající lidi. To je posíleno skutečností, že experimentální laboratoře bývají umístěny přímo v kampusu nebo budově univerzity a čas strávený cestou do ní představuje pro studenty (na rozdíl od většiny ostatních) jen minimální dodatečný náklad. Studenti se také ve srovnání s průměrnou populací rychleji učí a lépe porozumí i komplikovanějším experimentálním instrukcím. Je však zřejmé, že chování a rozhodování studentů univerzit může mít daleko k chování a rozhodování zbytku populace. Na tento problém upozornil Joseph Henrich a jeho kolegové (2010) v článku s názvem „*The weirdest people in the world?*“. Název článku je postaven na slovní hříčce se slovem WEIRD (divný). Toto slovo je v článku používáno ve smyslu akronymu složeného ze slov Western, Educated, Industrialized, Rich a Democratic. Článek tak poukazuje právě na skutečnost, že většina participantů laboratorních experimentů jsou vzdělaní lidé ze Západních, industrializovaných, bohatých a demokratických zemí. Tedy ze skupiny, která rozhodně není reprezentativní pro většinu světové populace. Výzkumné vzorky skládající se z participantů pocházejících z převážně amerických a evropských univerzit jsou proto často označováni jako **studentské populace**. Z metodologického pohledu tento problém však nehraje příliš velkou roli – ekonomická teorie totiž neurčuje vlastnosti populací, na kterých mají být teorie testovány. Protože není předepsán žádný „dobrý“ vzorek, nelze kritizovat výsledky experimentů pro využití „špatného“ vzorku (viz Boxík 19: Blame-the-theory argument). Kritika je však na místě v případě přílišné generalizace výsledků laboratorních výzkumů. Při psaní článků je proto nutné výzkumný vzorek charakterizovat a při interpretaci výsledků nahradit zobecňující výroky typu „*Lidé mají tendenci se rozhodovat tak a tak.*“ výroky typu „*Participantů této studie mají tendenci se rozhodovat tak a tak.*“

<sup>2</sup> To implicitně znamená, že čím více lidí je v databázi zaregistrováno, tím nižší show-up fee můžete vyplácet.

Dobrym příkladem, na kterém lze tuto problematiku ilustrovat, je Hra na ultimátum (Ultimatum Game). Ve Hře na ultimátum (představená Güth et al. (1982)) spolu hrají dva hráči – Navrhovatel a Příjemce. Navrhovatel na začátku hry obdrží částku  $e$  ECU a jeho úkolem je rozdělit tuto částku mezi sebe a Příjemce. Až do této chvíle je Hra na ultimátum stejná jako Hra na diktátora. Ve Hře na ultimátum má však Příjemce aktivní roli, protože musí učinit rozhodnutí, zda nabídku Navrhovatele přijme, nebo odmítne. Pokud nabídku přijme, hráči si částku rozdělí tak, jak určil Navrhovatel. Pokud odmítne, oba hráči obdrží nulu. Ve této hře Navrhovatelé typicky nabízí Příjemcům více než ve Hře na diktátora, a to v průměru 40 % částky. V průměru je 16 % nabídek odmítnuto (Oosterbeek et al., 2004). Částka, kterou Navrhovatelé nabízí, se však liší v závislosti na různých okolnostech. Jednou z hlavních okolností je výše prvotního vybavení, které Navrhovatel na začátku hry obdrží. Čím vyšší je prvotní vybavení, tím nižší jsou také nabídky Navrhovatelů (Andersen et al., 2011).

#### Boxík 19: Blame-the-theory argument

Blame-the-theory argument poukazuje na to, že pokud je experiment navržen k testování konkrétní teorie, pak jakékoliv jeho nerealistické prvky mohou stejně tak být považovány za nerealistické prvky teorie samotné. Tento argument bývá často používán k odmítnutí kritiky, že výsledky laboratorních experimentů mají nízkou externí validitu.

(Bardsley, 2010; Smith, 1982)

Henrich a kolegové v jiném článku (Henrich et al., 2005) porovnali tyto typické výsledky Hry na ultimátum (konkrétně z Pittsburghské studie od Roth et al. (1991)) s výsledky z 15 malých společností lovecko-sběračského typu. Na ose x bublinového grafu (Henrich et al., 2005) vidíte procentuální velikost nabídky<sup>3</sup>, šedivé horizontální sloupce značí průměrnou výši nabídky (podle které jsou jednotlivé společnosti na ose y seřazeny) a velikost jednotlivých bublin značí relativní četnost konkrétní nabídky (největší bublina je tedy modus – viz Boxík 20: Aritmetický průměr, medián a modus).

<sup>3</sup> Procenta vyjadřují podíl z celku, a proto jsou matematicky zapisována jako desetinné číslo mezi 0 a 1, například 0,2 nebo také  $\frac{2}{10}$  (zlomky jsou pohodlnější pro výpočty). Pokud toto desetinné číslo vynásobíme 100, získáme vyjádření, na které jsme zvyklí – 20 %.

**Boxík 20: Aritmetický průměr, medián a modus**

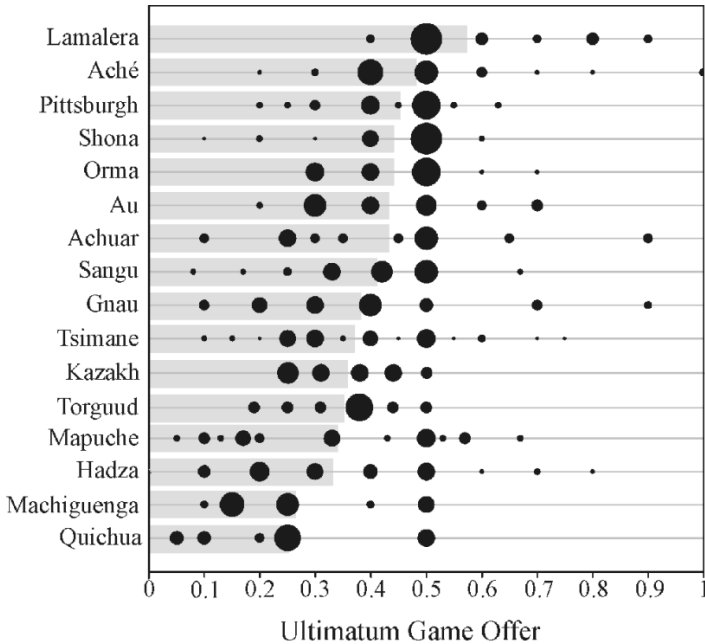
Představte si, že se participantů ptáte, jaký je jejich čistý příjem v tisících eur. Nasbíráte následujících sedm hodnot: 3, 1, 1, 10, 3, 1 a 9.

**Aritmetický průměr** (kterému se běžně říká jen „průměr“) těchto hodnot se vypočítá jako jejich součet vydělený jejich počtem. Tedy jako  $(3+1+1+10+3+1+9) / 7 = 4$ . Jak vidíte, průměr často bývá ovlivněn extrémními hodnotami, což vystihuje rčení „Když strčíte nohy do mrazáku a hlavu do trouby, v průměru je vám docela dobře.“

**Medián** je prostřední hodnota v souboru dat seřazeného vzestupně. V našem případě (1, 1, 1, 3, 3, 9, 10) je to tedy 3. Kdyby měl dataset sudý počet hodnot (například 1, 1, 1, 3, 3, 9), najdete prostřední dvě čísla (1 a 3) a jejich součet vydělíte dvěma. Medián by tedy byl  $(1+3) / 2 = 2$ . I když se často uvádí průměr nasbíraných hodnot, je vhodné uvádět i medián. V našem příkladu vypovídá o souboru medián 3 lépe než průměr 4, protože průměrného příjmu dosahují pouze dva lidé (s příjmem 9 a 10), zatímco mediánového čtyři lidé (s příjmem 3, 3, 9 a 10).

**Modus** pak je nejčastější hodnota v souboru dat, v našem případě 1.

Průměrná výše nabídky se v této studii pohybovala mezi 26 % (Quichua) a 58 % (Lamalera) s modem mezi 15 % (Machiguenga) a 50 %. Většina společností se tak pohybovala v rozmezí výsledků typických pro studentské populace, které v průměru nabízejí 40–50 % s modem 50 %. V grafu jsou však patrné mnohé výjimky. První z nich je vesnice Lamalera, kde průměrná výše nabídky jako jediná přesahuje 50 %. To je způsobeno především tím, že minimální výše nabídky je 40 % a nezanedbatelná část nabídek přesahuje 50 %. Na druhé straně stojí kmeny Machiguenga a Quichua s průměrnou výší nabídek 26 a 25 %, kde naopak nikdo nenabízí částky vyšší než 50 % a většina nabídek je nižších než 25 %. Jedno z vysvětlení těchto rozdílů, které autoři studie nabízí, je založeno na způsobu obživy. Zatímco Machiguenga a Quichua jsou převážně zemědělci, Lamaleřané jsou lovci velryb. Je zřejmé, že společenský tlak na spolupráci je vyšší ve společnostech, jejichž přežití na spolupráci přímo závisí (pro ulovení velryby je třeba spolupráce mnoha lidí) než ve společnostech, kde si obživu může zajistit jednotlivec sám (zemědělství). Možná vás nyní zajímá, proč jsou studentské populace svým chováním bližší lovcům velryb než zemědělcům. Západní společnosti jsou založené na spolupráci a důvěře, která je nezbytná pro fungování tržní ekonomiky. Nezapomeňte, že ve Hře na ultimátum má Příjemce možnost nabídku odmítnout, a tím Navrhovatele potrestat. Zdá se tedy, že čím více fungování společnosti závisí na kooperaci, tím více jsou členové této společnosti ochotni obětovat na potrestání těch, kteří nespolupracují dostatečně.



Obrázek 4: Procentuální výše nabídek při Hře na ultimátum  
Zdroj: Henrich et al. (2005)

Tento příklad tedy ukazuje, že výsledky experimentů prováděných na studentských populacích (nebo chcete-li WEIRD populacích) nejsou automaticky univerzálně aplikovatelné na všechny části společnosti. Jak jste mohli vyčíst mezi řádky, lidské chování a rozhodování je do značné míry ovlivněné kulturou dané společnosti (viz Boxík 21: Geneticko-kulturní koevoluce).

### Boxík 21: Geneticko-kulturní koevoluce

Geneticko-kulturní koevoluce (gene-culture coevolution) je koncept poukazující na skutečnost, že lidské chování je produktem dvou různých a vzájemně se ovlivňujících procesů: genetické evoluce a kulturní evoluce. Často uváděným příkladem je tolerance na laktózu. Všichni lidé jsou v kojeneckém věku schopni trávit mléko, avšak většina lidí o tuto schopnost přijde a konzumace mléčných produktů u nich vede k vážným zdravotním problémům. Výjimkou jsou potomci Evropanů (tedy i většina nynějších obyvatel Ameriky a Austrálie) a některých národů z oblasti severní Afriky, kteří domestikovali dobytek (slovo domestikace pochází z latinského *domus*, tedy dům – protože lidé až do zemědělské revoluce žili kočovným životem, měli bychom správně říkat, že dobytek domestikoval nás). Někdy během evoluční historie došlo u společného předka všech těchto populací k mutaci genu „vypínající“ produkci laktázy, která umožňuje trávení mléka. Tato mutace poskytla jemu a všem jeho potomkům evoluční výhodu, protože konzumaci mléka získali nejen dostupný a levný zdroj kalorií, ale také vápníku a vitamínu D, kterého měli lidé v severních zeměpisných šířkách nedostatku (je totiž získáván chemickou reakcí v kůži, která je spuštěna vystavením Slunci). Tím se zároveň vytvořil další evoluční tlak na chov dobytka, čímž se kruh geneticko-kulturní koevoluce uzavřel – kultura domestikace dobytka nejdřív umožnila přežít genetické mutace, jejíž rozšiřování dále podpořilo rozvoj této kultury.

(Dunbar & Barrett, 2007)

Má tedy smysl dělat experimentální výzkum na studentských populacích, i když výsledky nejsou reprezentativní? Ano, má! Nějak se přece začít musí – výsledky založené na datech od studentských populací slouží jako **baseline**. Ostatní výzkumníci pak mohou studii replikovat s použitím jiné populace a porovnat své výsledky s těmi původními. Tímto způsobem bylo například zjištěno, že studentské populace jsou v průměru méně prosociální než ostatní členové stejné společnosti (jedním z vysvětlení je tzv. *motivation crowding* – viz Boxík 22: *Motivation crowding*) (Falk et al., 2013). Jednou z důležitých implikací pro výzkumníky zabývající se experimentálním výzkumem společenských preferencí je, že studentské populace jim poskytují jakýsi minimální standard. To znamená, že pokud objeví nějaký efekt u relativně málo prosociální studentské populace, lze očekávat stejný (ne-li větší) efekt i u nestudentských populací<sup>4</sup>. Podobnou argumentační logiku lze samozřejmě využít ve všech oblastech experimentálního výzkumu, ve kterých byl nějaký minimální standard prokázán. Například Croson and Gneezy (2009) ukázali, že muži riskují více než ženy jak ve studentské populaci, tak v nestudentské populaci.

<sup>4</sup> Například Exadaktylos et al. (2013) ve své studii porovnávají výsledky studentských a nestudentských populací ve Hře na diktátora, Hře na Ultimátum a Hře na důvěru. Jediný rozdíl našli ve Hře na diktátora, kde se studentské populace liší od nestudentské maximálně o 6 %.

**Boxík 22: Motivation crowding**

Motivation crowding (volně přeložitelné jako vytěšňování motivace) je teorie poukazující na skutečnost, že odměňování určitého druhu chování může oslabit vnitřní motivaci k takovému chování. Například nabízení finanční odměny za účast v experimentech může odradit prosociální participanty, kteří by se experimentu zúčastnili pouze, pokud by měli pocit, že účastí pomáhají. Známý příklad se týká darování krve, kde nabízení finanční odměny způsobilo, že přestali chodit dlouhodobí a prověřeni dárci a začali chodit dárci, pro které byla i malá finanční odměna atraktivní (například bezdomovci nebo drogově závislí). Důvodem pro změnu chování dlouhodobých dárců bylo, že při existenci finanční odměny je darování krve pouhý obchod a nikoliv dobrý skutek. Jinými slovy, finanční odměny dlouhodobým dárcům nevyhradila ztrátu dobrého pocitu, který z darování krve do té doby měli.

(Mellström & Johannesson, 2008)

Tato kapitola poukázala na potenciální rizika pro externí validitu spojená se samovýběrem a využíváním studentských populací (WEIRD). Zároveň však ukazuje, že pokud si jsou výzkumníci těchto rizik vědomi, mohou je při interpretaci výsledků využít ve svůj prospěch. Vždy by však měli mít na paměti skutečnost, že experimentální výsledky jsou validní primárně pro výzkumný vzorek, na kterém byla studie provedena. Studentské vzorky však v žádném případě nejsou důvodem pro nerealizování experimentálních studií. Výzkumníci z jiných oborů často namítají, že by výzkum měl být prováděn na reprezentativním vzorku populace (viz Boxík 23: Reprezentativní vzorek). Zde je však třeba zapojit jednoduchou analýzu nákladů a přínosů. Reprezentativní vzorek se v prostředí Česka typicky skládá ze zhruba 1000 lidí. I když odhlédneme od skutečnosti, že by tento vzorek měl být složen z lidí všech sociodemografických skupin a především z odlišných regionů Česka, rekrutovat 1000 lidí pro účast v laboratorním experimentu by bylo extrémně nákladné a pro běžné výzkumníky nerealizovatelné. Přínosy českého reprezentativního vzorku jsou sporné již jen z toho pohledu, že experimentální ekonomie se snaží popisovat obecná pravidla rozhodování. Mohli bychom se pak ptát, zdali nejsou studenti z Česka podobnější studentům ve Spojených státech více, než je průměrný občan Česka podobný průměrnému občanovi Spojených států. Dalším argumentem proti používání reprezentativních vzorků v experimentální ekonomii je skutečnost, že experimenty mají ověřovat hypotézy. Oproti tomu například volební průzkumy mají mapovat realitu. Zatímco experiment tedy může skončit „neúspěchem“ (tj. důkazy o existenci hledaného efektu nebudou nalezeny), volební průzkum vždy skončí „úspěchem“ (tj. poskytne přehled o preferencích jednotlivých stran). Investovat proto obrovské množství úsilí, času a peněz do výzkumu, který by sice byl reprezentativní, ale může skončit „neúspěchem“, nedává žádný smysl. Mnohem racionálnější je ověřit hypotézy na menším, specifickém vzorku a v případě nevyvrácení existence zkoumaného efektu dále ověřovat, zdali lze stejný efekt najít i u jiných populací.

### 3.3 Určení velikosti výzkumného vzorku na základě velikosti efektu

Velikost výzkumného vzorku je jedním ze základních parametrů každé experimentální studie. O tom svědčí i skutečnost, že jedna z prvních informací, kterou se většinou dozvíte v sekci „Výsledky“ výzkumného článku (neboli „paperu“), je, kolik participantů se experimentu zúčastnilo. Zároveň je však těžké odpovědět na otázku, kolik participantů by se experimentu mělo zúčastnit. V ideálním světě, tedy ve světě s neomezeným rozpočtem a počtem lidí ochotných se studie zúčastnit, bychom velikost vzorku určili na základě vztahu **hladiny významnosti** (significance level), **statistické síly** (power) a **velikosti efektu** (effect size)<sup>5</sup>. K pochopení těchto jednotlivých konceptů je zapotřebí rozumět tomu, jak funguje výzkum založený na testování hypotéz.

#### Boxík 23: Reprezentativní vzorek

Reprezentativní vzorek je takový, ve kterém se charakteristika jednotlivců ve vzorku shoduje s charakteristikou celé populace. V reprezentativním vzorku tak musí být zastoupeni jednotlivci dle charakteristik jako pohlaví, věk, vzdělání, povolání, město bydliště, výše příjmu a tak dále. Velikost reprezentativního vzorku je v České republice typicky kolem 1000 lidí. Lidé mohou být do reprezentativního vzorku vybráni buď náhodně (například metodou CATI), nebo pomocí kvótního výběru.

Na začátku každého výzkumu stojí **výzkumná otázka**, na základě které je formulována jedna nebo více **hypotéz**. Zatímco výzkumná otázka může být relativně obecná (například, „Liší se výše částky zasílané Diktátorem ve Hře na Diktátora a Navrhovatelem ve Hře na Ultimátum?“), hypotéza musí být výrok, který je možný testovat. Každá hypotéza se pak skládá z **nulové hypotézy** (null hypothesis) a **alternativní hypotézy** (alternative hypothesis). Nulová hypotéza (označovaná jako  $H_0$ ) typicky konstatuje neexistenci efektu, který se snažíme najít:

*$H_0$ : Navrhovatelé nezasílají vyšší částky než Diktátoři.*

Alternativní hypotéza (označovaná jako  $H_1$ ), jak název napovídá, je opačný výrok konstatující existenci hledaného efektu:

*$H_1$ : Navrhovatelé zasílají vyšší částky než Diktátoři.*

<sup>5</sup> Do češtiny je tento pojem převážně překládán jako „velikost účinku“. My jsme se rozhodli pro „velikost efektu“ především z toho důvodu, že tento překlad lépe zapadá do terminologie experimentální ekonomie.



Následně uspořádáme experiment, ve kterém polovina participantů hraje Hru na Diktátora a polovina participantů hraje Hru na Ultimátum (participantů jsou do jedné nebo druhé hry přiřazeni náhodně a rozdělují stejně velké prvotní vybavení). Pokud provedená analýza odhalí, že průměrná částka zasílaná Navrhovatelem je **statisticky významně** (nebo také „signifikantně“) vyšší než průměrná částka zasílaná Diktátorem, znamená to, že můžeme konstatovat **nalezení podpory** pro naši alternativní hypotézu. V opačném případě – tedy pokud se zasílané částky v průměru neliší – konstatujeme nalezení podpory pro nulovou hypotézu.

V souvislosti s hypotézami se často můžete dočíst, že jedna nebo druhá hypotéza byla „zamítnuta“ nebo dokonce „potvrzena“. To je však chyba. Protože se našeho experimentu zúčastnil pouze nepatrný zlomek relevantní populace (v našem případě jí jsou všichni lidé), zjištěné výsledky jsou pouze aproximací reality, nikoliv jejím věrným obrazem. Vždy existuje šance, že námi zjištěný výsledek realitě neodpovídá, a proto maximum, co můžeme o našich zjištěních říci je, že podporují nulovou, nebo alternativní hypotézu.

V ideálním případě chceme, aby byla naše zjištění s realitou zarovnána. To znamená, že nalezneme podporu pro nulovou hypotézu, pokud zkoumaný efekt v realitě neexistuje, nebo podporu pro alternativní hypotézu, pokud zkoumaný efekt v realitě existuje. Může se však stát, že naše zjištění je **falešně pozitivní**. To znamená, že nalezneme podporu pro alternativní hypotézu i přestože efekt v realitě neexistuje. Druhou možností je, že naše zjištění je **falešně negativní**. To znamená, že nalezneme podporu pro nulovou hypotézu i přestože efekt v realitě existuje. Protože falešně pozitivní výsledek je pro výzkum potenciálně nebezpečnější (objevili jste něco, co v realitě neexistuje), označuje se ve statistice jako **Chyba prvního typu** (Type I. Error). Falešně negativní výsledek je pak spíše než pro výzkum problematický především pro výzkumníky (neobjevili jste něco, co v realitě existuje) a ve statistice se označuje jako **Chyba druhého typu** (Type II. Error).

O statistické významnosti výsledků (v našem případě rozdíl v průměrné výši zasílaných částek) rozhoduje, zdali jsme schopni s určenou minimální mírou pravděpodobnosti vyloučit, že jsme se dopustili obou chyb. V případě Chyby prvního typu se tato pravděpodobnost označuje **alfa** ( $\alpha$ ) a v případě Chyby druhého typu se označuje **beta** ( $\beta$ ) (viz Boxík 24: Řecká písmena). Pokud nám naše výsledky dávají vyšší jistotu než  $\alpha$  a  $\beta$  (resp. pravděpodobnost dopuštění se obou chyb je nižší než  $1-\alpha$  a  $1-\beta$ ), pak je můžeme považovat za statisticky významné. Tabulka 11 shrnuje popsané vztahy a následující dvě kapitoly je vysvětlují ve větším detailu.

		REALITA	
		H0 je pravdivá	H1 je pravdivá
VÝZKUM	Nalezení podpory pro H0	$1-\alpha$	Falešně negativní výsledek (Chyba druhého typu) $\beta$
	Nalezení podpory pro H1	Falešně pozitivní výsledek (Chyba prvního typu) $\alpha$	$1-\beta$

Tabulka 11: Možné situace při testování hypotéz

**Boxík 24: Řecká písmena** $\alpha$  alfa  
 $\beta$  beta $\Delta$  delta  
 $\theta$  théta $\iota$  ióta  
 $\mu$  mí $\sigma$  sigma (malá)  
 $\Sigma$  Sigma (velká)**3.3.1 Chyba prvního typu a hladina významnosti**

Hodnota  $\alpha$ , také označována jako **hladina významnosti** (significance level), je pravděpodobnost chyby prvního typu. Ve společenských vědách (kam patří i ekonomie) je hladina významnosti typicky stanovena na 0,05. To znamená, že jsme ochotni akceptovat maximálně 5% riziko, že výsledek statistické analýzy podporující alternativní hypotézu bude falešně pozitivní. V jiných odvětvích (například v medicíně či letectví) je hladina významnosti typicky mnohem nižší, často i méně než 0,001 (tj. 0,1 %). Představte si, že testujete účinky nového léku (H0: „Lék nefunguje.“, H1: „Lék funguje.“), který ve skutečnosti nefunguje (realita je tedy popsána nulovou hypotézou). Protože máte neomezené zdroje, udělali jste 100 experimentálních studií (každé studie se účastní jiné subjekty), ve kterých jste polovině pacientů dali placebo (kontrolní skupina) a polovině testovaný lék (experimentální skupina). Výsledkem je, že 97 studií (97 %) neprokázalo žádný rozdíl mezi skupinami a 3 studie (3 %) vlivem náhodného šumu rozdíl prokázaly. Nyní si však představte, že neomezené zdroje nemáte a můžete uspořádat pouze jednu z těchto studií. S trochou nadsázky si to lze představit tak, že výsledky oněch 100 virtuálních studií vhodíte do pytlíku, zatřepete s ním a následně si z něj vytáhnete pouze jeden výsledek. Pokud pracujete s hladinou významnosti 5 %, pak jste výsledek z tohoto pytlíku (ve kterém jsou jen 3 % falešně pozitivních výsledků) ochotni přijmout. Protože farmaceutické společnosti však nechtějí riskovat výrobu nefunkčního léku, budou ochotni akceptovat výsledky pouze z pytlíků, ve kterých bude méně než jeden falešně pozitivní výsledek na tisíc správných výsledků. Jinými slovy, budou ochotni přijmout pouze riziko menší než 0,1 %, že provedená studie poskytne falešně pozitivní výsledek.

Možná vás napadla otázka, jak určíte poměr falešně pozitivních výsledků v pytlíku, pokud oněch 100 studií opravdu neuskutečníte. Jinými slovy, jak určíte číslo, které budete porovnávat se stanovenou hladinou významnosti. Toto číslo se nazývá **p-hodnota** (p-value) a je výsledkem použitého statistického testu <sup>6</sup> (viz Boxík 25: Co je p-hodnota a co není?). Pokud je p-hodnota vyšší než zvolená hladina významnosti, pak konstatujeme nalezení podpory pro nulovou hypotézu, tedy že hledaný efekt neexistuje. V opačném případě, kdy je p-hodnota nižší než  $\alpha$ , konstatujeme nalezení podpory pro alternativní hypotézu, tedy že hledaný efekt existuje <sup>7</sup>. Pokud chcete o hladině významnosti a p-hodnotě přemýšlet méně abstraktně, představte si, že jsme nasbírali data do našeho experimentu a zjistili jsme, že Navrhovatelé zasílají v průměru 40 ECU, zatímco Diktátoři pouze 20 ECU. V případě, že zjištěná p-hodnota je nižší než 0,05, rozdíl mezi průměry lze považovat za statisticky významné (a tedy nalézáme podporu pro alternativní hypotézu). Pokud je p-hodnota vyšší než 0,05, rozdíl mezi průměry nelze považovat za statisticky významný (a tedy nalézáme podporu pro nulovou hypotézu).

Ještě zbývá doplnit, že k Chybě prvního typu často dochází kvůli problémům s externí validitou (výzkumný vzorek je zcela odlišný od zbytku populace), nebo interní validitou (reaktivita a nedostatečná randomizace).

#### Boxík 25: Co je p-hodnota a co není?

P-hodnota je pravděpodobnost získání výsledků, které jsou alespoň tak extrémní jako pozorované výsledky za předpokladu, že realita je popsána nulovou hypotézou. Jinými slovy, p-hodnota je pravděpodobnost, se kterou se nám podaří nasbírat naše data, pokud platí nulová hypotéza.

P-hodnota není pravděpodobnost, s jakou jsme získali výsledek dílem náhody, pravděpodobnost falešně pozitivního výsledku, ani pravděpodobnost, že nulová hypotéza je pravdivá. Nemá proto smysl mezi sebou porovnávat p-hodnoty různých testů a na základě toho usuzovat na „kvalitu“ zjištění. Obecně jde pouze říci, že v rámci jednoho testu je nižší p-hodnota lepší než vyšší p-hodnota.

<sup>6</sup> V tomto případě porovnáваме průměrné hodnoty ve dvou skupinách, použili bychom proto t-test.

<sup>7</sup> „When the P is low, the Ho must go!“ je mnemotechnická pomůcka využívající kombinace těchto vztahů, tedy že při nízké p-hodnotě musíme „zahodit“ nulovou hypotézu.

### 3.3.2 Chyba druhého typu a statistická síla

Hodnota  $\beta$  je pravděpodobnost Chyby druhého typu, tedy situace, kdy chybně konstatujeme podporu pro nulovou hypotézu na základě falešně negativního výsledku.  $\beta$  má typicky hodnotu 0,2 (tj. 20 %) <sup>8</sup>. Přímo souvisejícím konceptem pak je **statistická síla**, neboli pravděpodobnost správného konstatování podpory pro alternativní hypotézu. Statistická síla má hodnotu  $1-\beta$ , většinou tedy 0,8 (80 %).

Podobně jako v případě hladiny významnosti i o statistické síle lze přemýšlet následujícím způsobem. Neomezený rozpočet nám umožní stokrát uspořádat (stejnou) studii, která se snaží najít podporu pro alternativní hypotézu konstatující existenci vlastnosti, která se v populaci vyskytuje relativně málo – řekněme, že tuto vlastnost má v populaci 1 člověk ze 40. Protože však máme problém do našich sto studií sehnat participanty, nakonec se jich každé studie zúčastní pouze 10. Můžeme tedy očekávat, že jednoho participanta s hledanou vlastností objevíme v průměru v 1 studii ze 4, tedy ve 25 % studií (ponechme nyní stranou, zdali by výsledky založené na objevení této vlastnosti u 1 z 10 participantů byly statisticky významné na hladině  $\alpha$ ). Protože neomezený rozpočet (opět) nemáme, můžeme uspořádat pouze jednu z těchto studií. Znovu jsme tedy v situaci, kdy před námi leží pytlík s výsledky těchto virtuálních studií a my si z něj můžeme vylosovat jeden, na základě něhož budeme konstatovat podporu pro nulovou nebo alternativní hypotézu. V tomto případě pytlík obsahuje 25 % výsledků podporujících alternativní hypotézu a 75 % výsledků podporujících nulovou hypotézu. Pokud bychom z takového pytlíku byli ochotni losovat, pak se se 75% pravděpodobností dopustíme Chyby druhého typu, protože bychom našli podporu pro nulovou hypotézu o neexistenci hledané vlastnosti i přestože tato vlastnost ve skutečnosti existuje. Takové riziko však nebudeme ochotni podstoupit – „losování“ ve skutečnosti znamená přípravu a realizaci experimentální studie, která je časově i finančně nákladná. Aby pro nás bylo riziko neobjevení existujícího efektu přijatelné, budeme chtít vždy losovat z pytlíku, ve kterém je alespoň 80 % výsledků podporujících alternativní hypotézu.

Jak vyplývá z tohoto příkladu, Chyba druhého typu, resp. nedostatečná statistická síla, je typicky výsledkem kombinace malého vzorku a malé velikosti efektu. Detailně se těmito dvěma parametry věnuje následující část příručky.

<sup>8</sup> Protože snížení pravděpodobnosti Chyby druhého typu zvyšuje pravděpodobnost Chyby prvního typu (a naopak), je třeba zachovat vyváženost mezi  $\alpha$  a  $\beta$ . Obecně se má za to, že při  $\alpha = 0,05$  by měla být  $\beta = 0,2$  a při  $\alpha = 0,01$  by měla být  $\beta = 0,1$ .

### 3.3.3 Velikost efektu

**Velikost efektu** (effect size) je měřítko kvantifikující rozdíl mezi kontrolní a experimentální skupinou. V závislosti na typu dat se pro jeho výpočet využívají různé metody. Pro určení velikosti efektu při porovnávání průměrů se nejčastěji používá tzv. **Cohenovo d**. To se vypočítá jako rozdíl průměrů experimentální a kontrolní skupiny vydělený sdruženou směrodatnou odchylkou:

$$d = \frac{M_1 - M_2}{\frac{\sqrt{s_1^2 + s_2^2}}{2}}$$

Následující Tabulka 12 shrnuje relativní velikosti účinků pro vybrané hodnoty  $d$ . Z tabulky mj. vyplývá, že pokud je velikost účinku menší než 0,2, neměl by být rozdíl mezi průměrem kontrolní a experimentální skupiny přikládán velký význam, a to i pokud je tento rozdíl statisticky významný (tj.  $p$ -hodnota je nižší než 0,05). Naopak velikost účinku vyšší než 1 značí, že průměry kontrolní a experimentální skupiny se zásadně liší.

relativní velikost účinku	velikost účinku ( $d$ )	% kontrolní skupiny pod průměrem experimentální skupiny
	0,0	50 %
malá	0,2	58 %
střední	0,5	69 %
velká	0,8	79 %
	1,4	92 %

Tabulka 12: Relativní velikosti účinků pro vybrané hodnoty  $d$

Zdroj: Sullivan and Feinn (2012)

Pojďme si nyní koncept velikosti efektu vysvětlit na příkladu Hry na veřejné statky (Public goods game). Tu spolu hraje skupina hráčů, z nichž každý na začátku hry obdrží stejné prvotní vybavení ve výši  $e$  ECU. Úkolem každého hráče je rozhodnout se, jakou část svého prvotního vybavení vloží do veřejného fondu. Celková částka, která je všemi hráči do fondu vložena, je následně vynásobena předem určeným multiplikátorem  $m$  a rovnoměrně rozdělena mezi všechny členy skupiny

(včetně těch, kteří eventuálně nepřispěli nic)<sup>9</sup>. Participant v této hře typicky přispívají v průměru polovinu svého prvotního vybavení. V případě, že má hra více kol, však průměrný příspěvek klesá k 10 % (Chaudhuri, 2011) a hra tak směřuje do svého Nashova ekvilibria, kterým je černé pasažérství všech participantů. V posledním kole pak v průměru 73 % participantů nepřispívá nic (Fehr & Schmidt, 1999). Jedna z modifikací Hry na veřejné statky umožňuje participantům trestat na vlastní náklady ty členy skupiny, s jejichž vyšší příspěvku nejsou spokojeni. Výsledkem je, že až 80 % participantů přispívá ještě v posledním kole celé své prvotní vybavení (Fehr & Gächter, 2000)<sup>10</sup>.

Představte si, že chcete replikovat výsledky Hry na veřejné statky s možností trestání ve dvou zemích (Théta a Ióta), ve kterých takový výzkum ještě nebyl proveden. Protože máte křišťálovou kouli, víte, že strach z potrestání v Thétě motivuje ke spolupráci 9 lidí z 10 (prevalence 90 %), v Iótě pak pouze 1 člověka z 10 (prevalence 10 %). Ekonomické žurnály však nerady vidí publikace založené na okultismu a výsledky tak musíte získat experimentálně. Pokud bude mít kontrolní skupina v každé zemi 100 participantů, tak hledanou vlastnost objevíme v průměru u 90 théťanů a 10 ióťanů. Intuice nám napovídá, že na hladině významnosti 0,05 bychom mohli mít v Iótě problém najít podporu pro naši alternativní hypotézu. A to proto, že v našem vzorku zkrátka nebudeme mít dostatek participantů, kteří hledanou vlastnost mají. Naopak v Thétě bude participantů s touto vlastností dost téměř v jakkoliv velkém vzorku. Můžete o tom přemýšlet tak, že zatímco v Thétě je hledaný efekt v realitě velký (ve společnosti je velice běžný), v Iótě je hledaný efekt v realitě malý (ve společnosti je relativně vzácný).

Tuto intuici si můžeme ověřit na simulovaných datech<sup>11</sup>. Participant jsou v obou zemích náhodně zařazeni buď do kontrolní skupiny (Hra na veřejné statky bez potrestání), nebo do experimentální skupiny (Hra na veřejné statky s možností trestání). Průměrná výše příspěvku v kontrolní skupině je v obou zemích 40 ECU. Rozložení průměrné výše příspěvku v experimentálních skupinách odpovídá výše zmíněnému poměru. 90 % Théťanů a 10 % Ióťanů majících strach z potrestání přispělo v průměru 70 ECU. 10 % Théťanů a 90 % Ióťanů nemajících strach z potrestání přispělo v průměru 40 ECU<sup>12</sup>. Následující Tabulka 13 shrnuje výsledky t-testu a Cohenova d.

<sup>9</sup> Například každý z 10 hráčů této hry obdrží 100 ECU. Pokud 9 hráčů vloží do fondu celých 100 ECU a celková částka 900 ECU bude vynásobena multiplikátorem 2, pak fond bude mít velikost 1800 ECU. Ten pak bude rovnoměrně rozdělen mezi všech deset hráčů (každý tedy obdrží 180 ECU). Všimněte si, že desátý hráč (který si ponechal celé své prvotní vybavení) má na konci hry 280 ECU.

<sup>10</sup> Tyto výsledky ukazují (i přestože Hra na veřejné statky typicky postrádá kontext), jak by vypadal výběr daní, pokud by jejich placení bylo dobrovolné.

<sup>11</sup> Data byla získána z normálního rozdělení s průměry uvedenými dále v textu a se směrodatnou odchylkou  $s = 12$ .

<sup>12</sup> Pro zjednodušení předpokládáme, že participant nemající strach z potrestání se chovají stejně jako jejich kolegové v kontrolní skupině.

N = 100	průměr (kontrolní   experimentální)	p-hodnota (t-test)	podpořeno na hladině významnosti = 0,05	Cohenovo d
<b>Théta</b>	40   58	<.001	ANO	1,99 (= velký efekt)
<b>Ióta</b>	40   43	= .294	NE	0,21 (= malý efekt)

Tabulka 13: Výsledky t-testu a Cohenova d

Jak můžete vidět, při účasti 100 participantů (50 v kontrolní skupině a 50 v experimentální skupině) se podporu pro existenci efektu trestání podařilo nalézt v Thétě (p-hodnota < 0,001), avšak nikoliv v Iótě (p-hodnota = 0,294) (viz Boxík 26: Hladiny významnosti a zápis p-hodnot). Poslední sloupeček tabulky také potvrzuje naši úvahu o tom, že zatímco v Thétě je hledaný efekt velký (d = 1,22), v Iótě je malý (d = 0,21). Pokud bychom tedy byli výzkumníci v Iótě, zamítli bychom alternativní hypotézu i přestože efekt v realitě existuje, a tím bychom se dopustili chyby druhého typu. Jak bylo zmíněno výše, řešením tohoto problému je navýšení počtu participantů natolik, abychom byli schopni najít podporu pro existující, avšak relativně malý, efekt. Poslední otázkou tedy zůstává, jak najít „správný“ počet participantů.

**Boxík 26: Zápis p-hodnot**

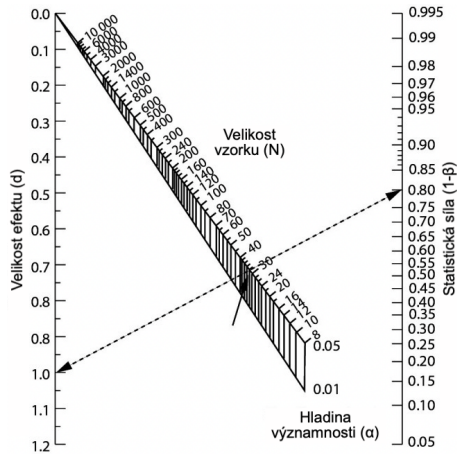
Nejčastěji uváděný výsledek většiny statistických testů je p-hodnota. Protože vyjadřuje pravděpodobnost (se kterou se nám podaří nasbírat naše data, pokud platí nulová hypotéza), její hodnota nabývá hodnot od 0 do 1. Podle citační normy Americké psychologické asociace (APA), která je v ekonomii využívána nejvíce, se p-hodnoty vyšší než 0.001 (= 0.1 %) zapisují zaokrouhlené na dvě nebo tři desetinná místa. Tato norma může být upravena žurnálem, ve kterém článek publikujete, avšak naše doporučení je využívat tři desetinná místa. Jelikož p-hodnota nemůže být vyšší než jedna, moderním pravidlem pro zápis p-hodnot je vynechávání úvodní nuly. Pokud je výsledkem testu p-hodnota 0.007989, zaokrouhlíte ji na tři desetinná místa (0.008) a zapíšete bez úvodní nuly, tedy „p-hodnota = .008“. Hodnoty nižší než 0.001 se pak zapisují jako „p-hodnota < .001“. Pokud jsou p-hodnoty opravdu malé, statistické softwary často volí tzv. vědecký zápis výsledku, například „1.011e-5“. Toto číslo lze jinak zapsat také jako „1.011 × 10<sup>-5</sup>“ a jeho převod do desetinného zápisu je vcelku jednoduchý – před samotné číslo 1.011 vložíte 5 nul (000001.011) a desetinnou tečku posunete za první nulu (0.00001011).

Pokud je p-hodnota nižší než zvolená hladina významnosti (ve společenských vědách typicky 0.05), a výsledky jsou tedy statisticky významné, pak se p-hodnoty mohou zapisovat také pomocí jedné až tří hvězdiček. Jedna hvězdička (\*) vyjadřuje p-hodnotu v rozmezí 0.05–0.01, dvě hvězdičky (\*\*) vyjadřují p-hodnotu v rozmezí 0.01–0.001 a tři hvězdičky (\*\*\*) vyjadřují p-hodnotu menší než 0.001. Někdy se také používá symbol (+), a to pro označení statisticky hraničně významného výsledku, tedy takového s p-hodnotou v rozmezí 0.1–0.05. Hvězdičky se typicky uvádí k číslu, jehož statistickou významnost značí. Pokud jsme tedy v Thétě naměřili průměry 40 a 58 a rozdíl (18) mezi kontrolní a experimentální skupinou je silně statisticky významný (p-hodnota < .001), pak tento výsledek můžeme zapsat také jako 18\*\*\*.

Poznámka: Protože se většina vědeckých článků píše v angličtině, použili jsme v tomto boxíku anglickou normu pro oddělovač desetinných míst, tedy tečku (.) místo v Česku standardní čárky (,).

Orientačně lze počet participantů odhadnout pomocí následujícího diagramu. Diagonála spojující velikost efektu (na levé vertikální ose) a požadovanou statistickou sílu (na pravé vertikální ose) protne škálu uprostřed diagramu. Tato škála v místě protnutí zobrazuje celkovou velikost vzorku pro hladinu významnosti 0,05 a 0,01. V modelovém příkladu, kde diagonála spojuje velikost efektu 1 (kontrolní a experimentální skupina se tedy v průměru liší o jednu směrodatnou odchylku) se statistickou silou 0,8, by se výzkumu mělo zúčastnit minimálně 32 participantů (16 v každé skupině) tak, abychom získali podporu pro hledaný efekt na hladině významnosti 0,05 (50 participantů v případě hladiny významnosti 0,01). V našem příkladu s velikostí efektu 0,21, požadovanou statistickou silou 0,8 a hladinou významnosti 0,05 by se v Ióté muselo výzkumu zúčastnit celkem zhruba 700 participantů (350 v každé skupině).





Obrázek 5: Diagram odhadu velikosti výzkumného vzorku. Přibližnou velikost výzkumného vzorku určuje průsečík číselné škály uprostřed diagramu a diagonály spojující velikost efektu (na levé vertikální ose) a požadovanou statistickou sílu (na pravé vertikální ose).

Převzato z: Serdar et al. (2021)

Pro přesné určení velikosti výzkumného vzorku pak můžeme použít následující vzoreček, ve kterém  $\Delta$  je absolutní hodnota rozdílu mezi průměry druhé a první výzkumné skupiny ( $|\mu_2 - \mu_1|$ );  $\sigma_1$  je rozptyl průměru první skupiny a  $\sigma_2$  je rozptyl průměru druhé skupiny;  $n_1$  je velikost vzorku první skupiny a  $n_2$  je velikost vzorku druhé skupiny;  $\alpha$  je hladina významnosti neboli pravděpodobnost Chyby prvního typu (většinou 0,05) a  $\beta$  je pravděpodobnost Chyby druhého typu (většinou 0,2); a konečně  $z$  je z-hodnota zvolených hodnot  $\alpha$  a  $\beta$  (1,96 pro  $\alpha = 0,05$  a 0,84 pro  $\beta = 0,2$ ).

$$n_1, n_2 = \frac{(\sigma_1^2 + \sigma_2^2) \left( z_{1-\frac{\alpha}{2}} + z_{1-\beta} \right)^2}{\Delta^2}$$

Abychom získali velikost vzorku, dosadíme do rovnice kromě z-hodnot požadované statistické síly (0,84 pro statistickou sílu 0,8) a hladiny významnosti (1,95 pro hladinu významnosti 0,05), také rozdíl průměrů experimentální (43 ze zadání) a kontrolní (40 ze zadání) skupiny a směrodatnou odchylku průměrů obou skupin (12 ze zadání).

$$n_1, n_2 = \frac{(12^2 + 12^2)(1,96 + 0,84)^2}{3^2} = 251$$

Po dosažení a výpočtu se dozvíme, že každá ze skupin by měla mít přesně 251 participantů. Výhodou je, že na internetu lze najít množství online kalkulačů, do kterých stačí pouze zadat hodnoty a výsledek se dozvíte bez nutnosti cokoli počítat.

Pokud bychom tedy studii v Iótě zopakovali s 500 participanty (počet jsme zaokrouhlili na desítky dolů), zjistíme, že p-hodnota se nyní rovná 0,013, a je tedy nižší než zvolená hladina významnosti. Můžeme proto konstatovat, že jsme – narozdíl od první studie se 100 participanty – našli podporu pro alternativní hypotézu. Výsledky všech testů jsou shrnuty v následující Tabulce 14.

	průměr (kontrolní   experimentální)	p-hodnota (t-test)	podpořeno na hladině významnosti = 0,05	Cohenovo d
<b>Théta</b> (N = 100)	40   58	<.001	ANO	1,99 (= velký efekt)
<b>Ióta</b> (N = 100)	40   43	= .294	NE	0,21 (= malý efekt)
<b>Ióta</b> (N = 500)	40   43	= .013	ANO	0,22 (= malý efekt)

Tabulka 14: Výsledky t-testu a Cohenova d po navýšení počtu participantů v Iótě

Tato kapitola mimo jiné poukazuje na význam **pilotních studií** v experimentální ekonomii. Pilotní studie lze rozdělit na dva typy. Prvním typem jsou technické pilotní studie, které jsou realizovány za účelem otestování funkčnosti experimentálního softwaru. Druhým typem jsou pilotní studie zaměřené na určení parametrů studie, které nelze předem vypočítat. Těmito parametry může být například průměrná výše variabilní odměny nebo právě údaje sloužící k odhadu velikosti efektu a výpočtu velikosti vzorku. V případě výzkumu v Iótě tak lze první výzkum považovat za pilotní studii, na základě které jsme byli schopni určit velikost vzorku, která nám umožní najít zkoumaný efekt.

### 3.4 Určení velikosti výzkumného vzorku na základě rozpočtu

Pokračujme nyní v příkladu z předchozí kapitoly, ve kterém jsme na základě pilotní studie se 100 participanty zjistili, že pro zjištění malého efektu bychom v Iótě potřebovali nasbírat data od dalších alespoň 400 participantů. Zde však můžeme narazit na dvě praktická omezení. Prvním omezením je, zdali jsme schopni prostřednictvím experimentální laboratoře takové množství participantů oslovit. Je třeba si uvědomit, že i pokud by se experimentu zúčastnilo 20 % z oslovených lidí, celkem by jich v databázi laboratoře muselo být 2500. Druhým, částečně souvisejícím, omezením je velikost rozpočtu, který máme pro realizaci studie k dispozici. Výše jsme zmínili, že čím větší procento zaregistrovaných chceme k účasti přimět, tím vyšší show-up fee musíme stanovit. I kdybychom tak počítali s průměrnou odměnou pouze ve výši 200 Kč, náklady na participanty by byly 100 tisíc korun. Alokovat takovou částku na odměny participantů může být náročný úkol nejen pro začínající experimentátory.

Pokud se dostaneme do situace, kdy z rozpočtu nejsme schopni uhradit optimálně vysoké odměny potřebnému počtu participantů, máme v zásadě dvě možnosti. První možností je snížit odměny na suboptimální úroveň a tím riskovat, že se nám nepodaří zaujmout tolik participantů, kolik potřebujeme. Pamatujte však, že experimentální laboratoře mají nastavenou minimální odměnu, za kterou jsou ochotny participanty na studie zvat. Druhou možností je snížit cílový počet participantů, zachovat optimální úroveň odměny a riskovat nedostatečnou statistickou silou a případný falešně negativní výsledek experimentu. V případě, že potřebujeme vyšší stovky participantů (tak jako v Iótě), pak nejspíš nastanou obě možnosti najednou, protože i při snížení odměny na minimální úroveň bude náš rozpočet stále nedostatečný a budeme tak muset snížit i cílový počet participantů. V takovém případě nám již nezbyvá nic jiného než doufat, že participanti i přes nízkou odměnu přijdou a nám se efekt podaří odhalit i s nižší statistickou silou.

Z této kapitoly mimo jiné vyplývá zásadní role pečlivého plánování experimentální studie již od jejího počátku. Především u začínajících výzkumníků se často objevuje snaha odpovědět v jedné studii na co nejvíce otázek, což se projeví na množství výzkumných skupin, které výsledný experimentální design obsahuje. To – v kombinaci s omezeným rozpočtem a chybějícím nebo nesprávným odhadem velikosti zkoumaných efektů – může vést k poddimenzované velikosti vzorku (underpowered). V důsledku toho pak studie není schopna odhalit efekty, které ve skutečnosti existují. Proto je při plánování studie nutné zohlednit nejen co bychom rádi zjistili, ale také výši rozpočtu a akviziční kapacity laboratoře, ve které bude sběr dat realizován.

## 3.5 Shrnutí

### Rekrutace participantů

Rekrutaci participantů většinou zajišťuje laboratoř, ve které experiment pořádáte. Počet přihlášených pak závisí na termínu studie a výši show-up fee. V závislosti na těchto faktorech se k účasti ve studii přihlásí typicky 10–20 % pozvaných.

### WEIRD participantů a reprezentativnost

Zobecnitelnost výsledků experimentálních studií často naráží na využívání tzv. WEIRD výzkumných vzorků. WEIRD je akronym pro **W**estern, **E**ducated, **I**ndustrialized, **R**ich a **D**emocratic a poukazuje na skutečnost, že většina participantů laboratorních experimentů je ze studentské populace, která není reprezentativní pro většinu světové populace. I přesto má smysl dělat experimentální výzkum na studentských populacích, jelikož slouží jako baseline. Experimentální studie prováděné na reprezentativním vzorku by byly extrémně nákladné a jejich organizace by vyžadovala početný výzkumný tým. Je proto racionálnější ověřit hypotézy na menším specifickém vzorku a v případě nevyvrácení existence zkoumaného efektu ověřovat stejný efekt i u jiných populací.

### Velikost výzkumného vzorku

Velikost výzkumného vzorku by měla být stanovena na základě vztahu hladiny významnosti, statistické síly a velikosti efektu. Na začátku výzkumu je formulována **výzkumná otázka**, na základě které je stanovena **hypotéza**. Hypotéza musí být výrok, který je možný testovat. Každá hypotéza má **nulovou hypotézu** typicky konstatující neexistenci efektu a **alternativní hypotézu** typicky konstatující existenci hledaného efektu. Na základě vyhodnocení statistického testu konstatujeme nalezení podpory pro jednu z hypotéz. Zjištěné výsledky statistického testu však nemusí odpovídat realitě – námi objevený výsledek může být falešně pozitivní (chyba prvního typu; alfa), nebo falešně negativní (chyba druhého typu, beta). **Falešně pozitivní výsledek** nastává, když odhalíme v realitě neexistující efekt. Míra rizika falešně pozitivního výsledku, kterou jsme ochotni akceptovat, se označuje jako **hladina významnosti** (alfa) a typicky bývá 0,05 (5 %). Pokud pravděpodobnost, se kterou se nám podaří nasbírat data podporující falešně pozitivní výsledek (p-hodnota), nepřesáhne zvolenou hladinu významnosti, pak konstatujeme nalezení podpory pro nulovou hypotézu. **Falešně negativní výsledek** nastává, když neodhalíme v realitě existující efekt. Pravděpodobnost chybného nalezení podpory pro nulovou hypotézu se označuje jako **beta** a typicky bývá 0,2 (20 %). Pravděpodobnost správného zamítnutí nulové hypotézy se nazývá **statistická síla**. Ta je vypočítána jako 1-beta a analogicky má hodnotu 0,8 (80 %). Nedostatečná statistická síla je

typicky výsledkem kombinace malého vzorku a malé **velikosti efektu**, což je měřítko kvantifikující rozdíl mezi kontrolní a experimentální skupinou. Obecně platí, že čím menší efekt se snažíme odhalit, tím větší výzkumný vzorek potřebujeme. V realitě však výzkumný vzorek spíše než velikostí hledaného efektu bývá určen velikostí rozpočtu, kterým výzkumníci disponují.



# Literatura

- Andersen, S., Ertac, S., Gneezy, U., Hoffman, M., & List, J. A. (2011). Stakes matter in ultimatum games. *American Economic Review*, 101(7), 3427–3439. <https://doi.org/10.1257/aer.101.7.3427>
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. *Groups, leadership, and men*, 177–190.
- Bardsley, N. (2010). Sociality and external validity in experimental economics. *Mind & Society*, 9(2), 119–138. <https://doi.org/10.1007/s11299-010-0075-0>
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, 10(1), 122–142. <https://doi.org/10.1006/game.1995.1027>
- Berger, L., & Bosetti, V. (2020). Characterizing ambiguity attitudes using model uncertainty. *Journal of Economic Behavior & Organization*, 180, 621–637. <https://doi.org/10.1016/j.jebo.2020.02.014>
- Camerer, C. F. (1997). Progress in behavioral game theory. *Journal of economic perspectives*, 11(4), 167–188. <https://doi.org/10.1257/jep.11.4.167>
- Camerer, C. F. (1999). Behavioral economics: Reunifying psychology and economics. *Proceedings of the National Academy of Sciences*, 96(19), 10575–10577. <https://doi.org/10.1073/pnas.96.19.10575>
- Camerer, C. F. (2011). The promise and success of lab-field generalizability in experimental economics: A critical reply to levitt and list. *Available at SSRN 1977749*. <https://doi.org/10.2139/ssrn.1977749>
- Cárdenas, J. C., De Roux, N., Jaramillo, C. R., & Martinez, L. R. (2014). Is it my money or not? an experiment on risk aversion and the house-money effect. *Experimental Economics*, 17, 47–60. <https://doi.org/10.1007/s10683-013-9356-x>
- Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, 149, 74–87. <https://doi.org/10.1016/j.jebo.2018.02.024>
- Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental economics*, 14, 47–83. <https://doi.org/10.1007/s10683-010-9257-1>
- Cherry, T. L., Frykblom, P., & Shogren, J. F. (2002). Hardnose the dictator. *American Economic Review*, 92(4), 1218–1221. <https://doi.org/10.1257/00028280260344740>
- Colburn, T., & Shute, G. (2007). Abstraction in computer science. *Minds and Machines*, 17, 169–184. <https://doi.org/10.1007/s11023-007-9061-7>
- Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, 47(2), 448–474. <https://doi.org/10.1257/jel.47.2.448>

- Cueva, C., Roberts, R. E., Spencer, T., Rani, N., Tempest, M., Tobler, P. N., Herbert, J., & Rustichini, A. (2015). Cortisol and testosterone increase financial risk taking and may destabilize markets. *Scientific reports*, 5(1), 1–16. <https://doi.org/10.1038/srep11206>
- Damasio, A. R. (1994). Descartes' error and the future of human life. *Scientific American*, 271(4), 144–144. <https://doi.org/10.1038/scientificamerican1094-144>
- Dickerson, C. A., Thibodeau, R., Aronson, E., & Miller, D. (1992). Using cognitive dissonance to encourage water conservation 1. *Journal of applied social psychology*, 22(11), 841–854. <https://doi.org/10.1111/j.1559-1816.1992.tb00928.x>
- Dunbar, R. I. M., & Barrett, L. (2007). *Oxford handbook of evolutionary psychology*. Oxford University Press, USA.
- Engel, C. (2011). Dictator games: A meta study. *Experimental economics*, 14, 583–610. <https://doi.org/10.1007/s10683-011-9283-7>
- Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological science*, 12(5), 391–396. <https://doi.org/10.1111/1467-9280.00372>
- Exadaktylos, F., Espin, A. M., & Branäs-Garza, P. (2013). Experimental subjects are not different. *Scientific reports*, 3(1), 1213. <https://doi.org/10.1038/srep01213>
- Falk, A., Meier, S., & Zehnder, C. (2013). Do lab experiments misrepresent social preferences? the case of self-selected student samples. *Journal of the European Economic Association*, 11(4), 839–852. <https://doi.org/10.1111/jeea.12019>
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4), 980–994. <https://doi.org/10.1257/aer.90.4.980>
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3), 817–868. <https://doi.org/10.1162/003355399556151>
- Feltovich, N. (2011). What's to know about laboratory experimentation in economics? *Journal of Economic Surveys*, 25(2), 371–379. <https://doi.org/10.1111/j.1467-6419.2010.00676.x>
- Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic behavior*, 6(3), 347–369. <https://doi.org/10.1006/game.1994.1021>
- Friedman, M., & Savage, L. J. (1952). The expected-utility hypothesis and the measurability of utility. *Journal of Political Economy*, 60(6), 463–474. <https://doi.org/10.1086/257308>
- Gelo, O., Braakmann, D., & Benetka, G. (2009). Quantitative and qualitative research: Beyond the debate. *Integrative psychological & behavioral science*, 43(4), 406. <https://doi.org/10.1007/s12124-008-9078-3>



- Gjerstad, S. D., & Smith, V. L. (2014). *Rethinking housing bubbles: The role of household and bank balance sheets in modeling economic cycles*. Cambridge University Press.
- Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization*, 3(4), 367–388. [https://doi.org/10.1016/0167-2681\(82\)90011-7](https://doi.org/10.1016/0167-2681(82)90011-7)
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., et al. (2005). “economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and brain sciences*, 28(6), 795–815. <https://doi.org/10.1017/S0140525X05000142>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences*, 33(2-3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, 92(5), 1644–1655. <https://doi.org/10.1257/000282802762024700>
- Holt, C. A., & Roth, A. E. (2004). The nash equilibrium: A perspective. *Proceedings of the National Academy of Sciences*, 101(12), 3999–4002. <https://doi.org/10.1073/pnas.0308738101>
- Hurwicz, L. (1972). On informationally decentralized systems. *Decision and organization: A volume in Honor of J. Marschak*.
- Iyengar, S., & Kinder, D. R. (1987). *News that matters: Television and american opinion*. University of Chicago Press.
- Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of economic psychology*, 32(5), 865–889. <https://doi.org/10.1016/j.joep.2011.05.007>
- Kahneman, D. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47, 278. <https://doi.org/10.2307/1914185>
- Kahneman, D. (2011). *Thinking, fast and slow*. macmillan.
- Kramer, J. (2007). Is abstraction the key to computing? *Communications of the ACM*, 50(4), 36–42. <https://doi.org/10.1145/1232743.1232745>
- Meeker, D., Linder, J. A., Fox, C. R., Friedberg, M. W., Persell, S. D., Goldstein, N. J., Knight, T. K., Hay, J. W., & Doctor, J. N. (2016). Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: A randomized clinical trial. *Jama*, 315(6), 562–570. <https://doi.org/10.1001/jama.2016.0275>
- Mellström, C., & Johannesson, M. (2008). Crowding out in blood donation: Was titmuss right? *Journal of the European Economic Association*, 6(4), 845–863. <https://doi.org/10.1162/jeea.2008.6.4.845>
- Oosterbeek, H., Sloof, R., & Van De Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental economics*, 7, 171–188. <https://doi.org/10.1023/b:exec.0000026978.14316.74>

- Ortmann, A. (2005). Field experiments in economics: Some methodological caveats. In *Field experiments in economics* (pp. 51–70). Emerald Group Publishing Limited. [https://doi.org/10.1016/S0193-2306\(04\)10003-3](https://doi.org/10.1016/S0193-2306(04)10003-3)
- Oswald, M. E., & Grosjean, S. (2004). *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory*. In R. F. Pohl (ed.) Psychology Press.
- Payne, J. G. (2010). The Bradley effect: Mediated reality of race and politics in the 2008 US presidential election. *American Behavioral Scientist*, 54(4), 417–435. <https://doi.org/10.1177/0002764210381713>
- Posner, R. A. (1997). Rational choice, behavioral economics, and the law. *StAn. l. rev.*, 50, 1551. <https://doi.org/10.2307/1229305>
- Roe, B. E., & Just, D. R. (2009). Internal and external validity in economics research: Tradeoffs between experiments, field experiments, natural experiments, and field data. *American Journal of Agricultural Economics*, 91(5), 1266–1271. <https://www.jstor.org/stable/20616293>
- Ronay, R., & Hippel, W. v. (2010). The presence of an attractive woman elevates testosterone and physical risk taking in young men. *Social Psychological and Personality Science*, 1(1), 57–64. <https://doi.org/10.1177/1948550609352807>
- Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *The American Economic Review*, 1068–1095. <https://www.jstor.org/stable/2006907>
- Saretsky, G. (1972). The OEO PC experiment and the John Henry effect. *The Phi Delta Kappan*, 53(9), 579–581. <https://www.jstor.org/stable/20373317>
- Serdar, C. C., Cihan, M., Yücel, D., & Serdar, M. A. (2021). Sample size, power and effect size revisited: Simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia medica*, 31(1), 27–53. <https://doi.org/10.11613/bm.2021.010502>
- Simon, H. A. (1990). Bounded rationality. *Utility and probability*, 15–18. [https://doi.org/10.1007/978-1-349-20568-4\\_5](https://doi.org/10.1007/978-1-349-20568-4_5)
- Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy*, 70(2), 111–137. <https://doi.org/10.1086/258609>
- Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review*, 66(2), 274–279. <https://www.jstor.org/stable/1817233>
- Smith, V. L. (1982). Microeconomic systems as an experimental science. *The American Economic Review*, 72(5), 923–955. <https://www.jstor.org/stable/1812014>
- Smith, V. L. (2018). *A life of experimental economics, volume i: Forty years of discovery*. Springer.
- Smith, V. L., Suchanek, G. L., & Williams, A. W. (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica: Journal of the Econometric Society*, 1119–1151. <https://doi.org/10.2307/1911361>

- Sonnemans, J., & Offerman, T. (2001). Is the quadratic scoring rule really incentive compatible?
- Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communication: Social determinants of information use in judgments of life satisfaction. *European journal of social psychology*, 18(5), 429–442. <https://doi.org/10.1002/ejsp.2420180505>
- Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the p value is not enough. *Journal of graduate medical education*, 4(3), 279–282. <https://doi.org/10.4300/jgme-d-12-00156.1>
- Vigen, T. (2022). *Spurious correlations*. Retrieved October 30, 2022, from <https://tylervigen.com/spurious-correlations>
- Weber, S. J., & Cook, T. D. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin*, 77(4), 273. <https://doi.org/10.1037/h0032351>
- Wickström, G., & Bendix, T. (2000). The "hawthorne effect"—what did the original hawthorne studies actually show? *Scandinavian journal of work, environment & health*, 363–367. <https://www.jstor.org/stable/40967074>





ISBN 978-80-7561-410-0



9 788075 614100