

7 Metodologie statistických výzkumů

Organizace statistických šetření, problematika náhodných výběrů, tvorba dotazníků, reliabilita a validita dotazníků, atd. Časté chyby při statistickém zpracování dat.

Obsah

7.1	Kvantitativní výzkum	2
	Výhody kvantitativního výzkumu	3
	Nevýhody kvantitativního výzkumu	3
7.2	Výzkumný problém/otázka	4
	7.2.1 Cíle	6
	7.2.2 Hypotézy	6
7.3	Metody sběru dat	8
	7.3.1 Pravděpodobnostní výběry (Disman, 2002)	8
	7.3.2 Záměrný výběr (Hendl, 2005, 2006)	10
	Kvótní Výběr	10
	Účelový výběr	10
	Výběr na základě dobrovolnosti	10
	Výběr na základě dostupnosti	11
	Technika sněhové koule	11
7.4	Tvorba dotazníku/testů	12
	7.4.1 Zásady tvorby dotazníku:	12
	7.4.2 Typy položek v dotazníku	13
	Uzavřené	13
	Otevřené	15
	Polouzavřené	16
	Filtrovní	16
	Projekční	16
	Kontrolní	17
	7.4.3 Závěrem k tvorbě dotazníku	17

7.4.4 Validita dotazníku	17
Obsahová validita	18
Konstruktová validita	19
Kriteriální validita	20
Faktory ovlivňující validitu	21
Test retest reliabilita	22
Reliabilita paralelních měření	22
Reliabilita testu jako vnitřní konzistence	23
7.5 Časté chyby při statistickém zpracování dat	26
7.5.1 Chyba prvního a druhého druhu	26
7.5.2 Korelace vs. kauzalita	27
7.5.3 Chyba měření	27

Kvantitativní výzkum – výzkumný problém – metody sběru dat – tvorba dotazníku – validita a reliabilita

7.1 Kvantitativní výzkum

Výzkumem se rozumí systematická tvůrčí práce rozšiřující poznání o nejzákladnějších příčinách jevů a pozorovatelných skutečností, včetně poznání člověka, kultury nebo společnosti, metodami umožňujícími potvrzení, doplnění či vyvrácení získaných poznatků. Je prováděn v zájmu rozvoje poznání, bez snahy o aplikaci výsledků na řešení praktických poměrů (Hendl, 2012, s. 24). Výzkum dle této definice se někdy též označuje jako **výzkum základní** (typicky např. výzkum ve fyzice) v protikladu k **výzkumu aplikovanému**, jehož cílem jsou aplikace poznatků základního výzkumu do praxe (typicky např. výzkum v elektrotechnice).

Dle použité metodologie lze rozdělit výzkum na **kvalitativní a kvantitativní**. Oba přístupy v sobě zahrnují různé výzkumné metody, nicméně jejich použití dává náhled na celkový obraz sledované problematiky. „*První se pokouší vysvětlit lidské chování skrz příčiny a následky, zatímco ten druhý má za cíl porozumět a interpretovat lidské aktivity skrze realitu každého individuálně*“ (Finn et al., 2000, s.37).

Kvantitativní výzkum (objektivní paradigma) je převážně založen na sběru dat, která jsou pak analyzována různými statistickými technikami. Směřuje k získání relativně malého množství informací od velkého počtu respondentů. Zahrnuje precizní proces hypotetické

formulace, detašované pozorování, sběr dat, analýzu dat a akceptaci, nebo odmítnutí hypotézy (Weaver a Lawton, 2002).

Jedním z typů kvantitativního výzkumu je **experimentální výzkum**, v němž manipulujeme s vybraným vzorkem a pak měříme dopady (obvyklá metoda výzkumu v přírodních vědách, v medicíně tzv. klinický výzkum). Druhým typem je **šetření** (observace), kdy výzkumník aktivně realitu neovlivňuje, je pouze jejím pozorovatelem (např. statistické šetření, v medicíně tzv. klinické studie).

Výhody kvantitativního výzkumu

- Pokud zvolíte tuto metodu (dotazník pro kvantitativní výzkum), přinese Vám velké množství responsí – odpovědí – od klientů, zákazníků, uživatelů a ostatních skupin obyvatel, na které jste se Vy nebo Vaše organizace zaměřili. Na základě statistického zpracování kvantitativně získaných dat můžete poznatky využívat k efektivnějšímu rozhodování, přesnějšímu plánování, komunikaci se zákazníky apod.
- Tato metoda se také vyznačuje tím, že je rychle proveditelná, levná a zvládnutelná jednotlivci, zejména je-li použit online dotazník ke sběru dat.

Nevýhody kvantitativního výzkumu

- Výsledky z kvantitativního výzkumu mohou být příliš obecné. Ne vždy totiž dokážou popsat problém do hloubky.
- Výzkumník může opomenout důležité vlastnosti zkoumaného vzorku, protože se soustředí na konkrétní problém a nemusí počítat s širší oblastí problému. (Hendl, 2012).

Základní rozdíly kvantitativního a kvalitativního výzkumu (specificky ve společenských vědách) je možné shrnout do následující tabulky:

Kvantitativní sběr dat	Kvalitativní sběr dat
Výzkumný vzorek lidí je velký počet respondentů	Výzkumný vzorek lidí je malý počet respondentů
Provádí se především pomocí dotazníkových šetření	Provádí se především pomocí osobních rozhovorů
Zkoumá problémy okrajově	Zkoumá problémy do hloubky
Časově nenáročný	Časově náročný
Dedukce z výsledků	Indukce z výsledků
Statistické zpracování dat	Nestatistické zpracování dat

Zdroj: Walker (2013)

Kvantitativní výzkum je tedy zaměřen na kvantitu, což je vlastnost, která je měřitelná a vyjádřitelná číslem.

Charakteristickými rysy kvantitativního výzkumu (opět specificky ve společenských vědách) podle Pavlici a kol. (2000, s. 27-28) jsou:

- 1) **nezávislost** – výzkumník je nezávislý na zkoumaných jevech;
- 2) **hodnotová svoboda a autonomie vědy** – výběr a volba toho, co a jakým způsobem bude studováno, by měly být determinovány objektivními kritérii (např. výsledky předchozích výzkumů, poukazující na určité problémy);
- 3) **kauzalita** – cílem sociální vědy by měla být identifikace kauzálních vztahů a zákonitostí, které objasňují pravidelnost lidských projevů;
- 4) **hypoteticko-deduktivní přístup** – věda se rozvíjí prostřednictvím procesu formulace a testování hypotéz, týkajících se obecných pravidel a zákonitostí;
- 5) **operacionalizace** – vědecké pojmy by měly být operacionalizovány (tj. převedeny do řeči konkrétních projevů a faktů), aby tak bylo umožněno kvantitativní měření skutečností, ke kterým se vztahují a které charakterizují;
- 6) **redukcionismus** – problémům lze celkově lépe porozumět tehdy, jsou-li redukovány na co nejjednodušší elementy (s tímto předpokladem však všichni pozitivisté nesouhlasí);
- 7) **generalizace** – možnost zobecnění zjištěných zákonitostí lidského a sociálního chování je nutno zajistit především dostatečně velkými vzorky zkoumaných osob;
- 8) **průřezová analýza** – obecná pravidla a zákonitosti lze nejlépe identifikovat prostřednictvím srovnávání variací napříč různými vzorky

Poznámka: Jedná se o pojmy, s kterými by se student měl seznámit dříve, než bude interpretovat své závěry.

7.2 Výzkumný problém/otázka

Každý z výzkumů by měl začít formulací výzkumného problému. Zvolit a adekvátně naformulovat výzkumný problém není možné bez pročtení dostatečného množství zdrojů. Sednout si a vymyslet vhodný výzkumný problém „z paty“ není nejvhodnějším způsobem. Například Jeřábek (1992) popisuje šest kroků od výzkumného záměru (toho, co asi chci dělat) k definici výzkumného problému (zde již žádné asi není).

- seznámení se známými výsledky výzkumů ve zkoumané oblasti,
- ohraničení problematiky,
- využití ověřených postupů řešení,
- vyvarování se chyb předchozích výzkumů,
- samostatné řešení vyvstalé otázky,
- formulace otázky nové

Platí, že výzkumným problémem je tázací věta, která musí splňovat hned několik předpokladů:

- a) Zpravidla na ni nelze odpovědět ANO-NE
Ukázka chybné formulace: Je možné pozitivně rozvíjet prostorovou orientaci žáků?
- b) Musí být ověřitelná
Ukázka chybné formulace: Používají hráči NFL doping častěji, než cyklisti? Tuto skutečnost nelze ověřit.
- c) Musí být hodnotná a mít smysl
Ukázka chybné formulace: Pomáhá domácí učení rozvoji žáka? Je jasné, že ano a tak není co ověřovat.
- d) Nesmí být příliš obecná
Ukázka chybné formulace: Jak má vypadat správné vyučování matematice? Zde není jasné, co a jak se bude zkoumat.

Po vymezení problému je třeba **definovat hlavní pojmy**. K definici pojmů se využívají slovníky a encyklopedie (pro pedagogiku např. Průcha, Walterová, Mareš: Pedagogický slovník, 2001). Pojmy musíme definovat tak, aby byly zjistitelné, resp. měřitelné¹.

Existuje hned několik druhů výzkumných problémů:

1. Deskriptivní

Zjišťuje a popisuje aktuální stav/situaci. Jedná se o jeden z nejjednodušších výzkumných problémů, jelikož pro jeho ověření zpravidla stačí základní deskriptivní analýza.

Ukázka: ***Jaké jsou fyzické proporce sprinterů startujících na olympiádě?***

2. Relační

Dáváme do vztahu dvě a více proměnných. Ptáme se, zda existuje vztah mezi zkoumanými jevy a jak je vztah těsný. Základní charakteristikou je, že pracujeme s jednou skupinou osob.

Ukázka: ***Jaký je vztah mezi motivací žáka a prospěchem z matematiky?***

3. Kauzální

Již podle názvu je zřejmé, že se jedná o výzkumný problém, který zjišťuje nějakou příčinu zkoumaného jevu. Tento typ výzkumného problému je charakteristický tím, že pracujeme s dvěma a více skupinami osob.

Ukázka: ***Jaký je vliv pohlaví na startovní reakci vycházející z akustického podnětu?***

¹ S tímto souvisí tzv. operativní definice pojmu

Ve chvíli, kdy je naformulovaný výzkumný problém je vhodné tento problém doplnit o sérii cílů. Tyto cíle ve své podstatě plní funkce jakéhosi návodu, který vám říká, co vše je nutné zvládnout k tomu, aby bylo možné ověřit výzkumný problém.

7.2.1 Cíle

Jsou-li cíle výzkumu jasně definovány, je to významný základ efektivity a úspěšnosti další činnosti. Tyto cíle tvoří jakýsi návod k tomu, jak nadále pokračovat. Cílů známe hned několik

- **orientační** – slouží k doplnění poznatků (např. Zjistit, jak jsou vnímáni svým okolím muži na rodičovské dovolené)
- **deskriptivní** – slouží k popsání stavu (např. Zjistit, z jakých zdrojů plyne hlavní příjem žen na mateřské dovolené),
- **explorační** – slouží k vysvětlení souvislostí (např. Zjistit, jaký vliv má množství volného času osob na jejich čas strávený sportem?),
- **sociotechnický** – slouží k formulaci návrhů, optimalizaci postupů (např. Zjistit, jaký model chování manažera zaměstnancům nejvíce vyhovuje).

Ve chvíli, kdy je naformulovaný výzkumný problém doplněný o dostatečný počet cílů, je vhodné vymezit hypotézy a to zejména z toho důvodu, že výzkumný problém je poměrně široký/obecný, kdežto hypotézy již hovoří o konkrétním „zkoumatelném“ jevu. Ve své podstatě se jedná o rozdělení výzkumného problému do dílčích, ověřitelných částí, které je možné buď potvrdit, nebo vyvrátit.

7.2.2 Hypotézy

Hypotéza² je vědecký předpoklad hovořící o vztahu dvou a více proměnných. Je důležité si uvědomit, že hypotéza řídí výzkum, nelze tudíž začít sběrem dat a během něj či dokonce po něm pak vymýšlet hypotézy! Při formulaci hypotéz často dochází k chybám, je proto důležité dodržovat tři následující pravidla:

1. Hypotéza hovoří o vztahu dvou a více proměnných³.
2. Hypotéza je formulována jako oznamovací věta.
3. Hypotéza musí být testovatelná.

² V tomto případě hovoříme o statistických hypotézách. Někdy se formulují „pracovní“ hypotézy, které nejsou totožné se statistickými hypotézami.

³ Zpravidla tomu tak je. Je možné, aby hypotéza pracovala jen s jednou proměnnou.

Ukázka chybných formulací:

Nejčastější nedostatky:

- Hypotéza dává do vztahu proměnné, u kterých není jasné, že daný vztah vůbec existuje.
- Vlastní formulace hypotézy je příliš náročná a zdlouhavá. Hypotézy je lepší formulovat jasně, stručně, věcně a výstižně.
- Student tvoří více hypotéz, ale není jasné proč, z jakého důvodu a jaký je mezi nimi vztah. Mělo by vždy být možné formulovat jednu hypotézu, která tyto „dílní“ hypotézy zastřešuje.
- Nemožnost potvrzení hypotézy.

Hypotézy je možné dělit na věcné a nulové. Věcné hypotézy jsou pracovní a ty neověřujeme. Je vždy nutné je přeformulovat jako nulové (statistické), které se následně ověřují pomocí vhodných statistických metod.

Ukázka:

Věcná hypotéza: Startovní reakce mužů na akustický podnět je rychlejší, než startovní reakce žen.

Nulová hypotéza: Není rozdíl ve startovní reakci mužů a žen.

S pojmem hypotéza souvisí také pojem operacionalizace, který říká, že vědecké pojmy by měly být operacionalizovány (tj. převedeny do řeči konkrétních projevů a faktů), aby tak bylo umožněno kvantitativní měření skutečností, ke kterým se vztahují a které charakterizují (Pavlica, 2000, s. 27).

Popis jednotlivých kroků konstrukce hypotézy popisuje Vojtíšek (2012).



Obecná hypotéza: první nástřel hypotézy vycházející z předpokladu existence závislosti.

Operacionalizace: získání konkrétních proměnných z pojmů, s kterými jsme doposud pracovali v obecné rovině

Pracovní hypotéza: Zde se již jedná o hypotézu vycházející z předpokladu závislosti dvou a více proměnných. Tyto proměnné jsou získány na základě operacionalizace.

Vztah těchto pojmů uvádí Vojtíšek (2012) v jednoduché tabulce:

Obecná hypotéza	Proměnné s validní operační definicí	Pracovní hypotéza
„Studijní výsledky studentů VOŠ závisejí na ekonomickém zázemí jejich rodiny.“	<p>- <i>závislá proměnná</i>: studijní výsledky měřené známkovým průměrem (1-3)</p> <p>- <i>nezávislá proměnná</i>: kvalita ekonomického zázemí měřená čistým příjmem rodiny v intervalech např.: do 15.000 Kč, 15.000-25.000 Kč, 25.000-30.000 Kč, více než 35.000 Kč</p> <p>- studenti VOŠ: 1., 2., 3. ročníku, obor studia</p>	<p>H1: Studenti VOŠSP z rodin s čistým příjmem nad 35.000 Kč dosahují v průměru nejlepších studijních výsledků</p> <p>H2: Studenti VOŠSP z rodin s čistým příjmem pod 15.000 Kč dosahují v průměru nejhorších studijních výsledků.</p>

7.3 Metody sběru dat

Častým problémem statistického šetření je samotná reprezentativnost naměřených dat. Je zřejmé, že v rámci výzkumu není možné postihnout celou zkoumanou realitu⁴ (například pracovat se všemi žáky základních škol), ale pouze s jistým výběrem. V tomto případě hovoříme o **reprezentativním** nebo **výběrovém**⁵ souboru. Tento reprezentativní/výběrový soubor musí imitovat složení populace tak přesně, jak je to jen možné⁶, s rostoucí velikostí vzorku se rozdíl mezi strukturou populace a vzorku zmenšuje. V krátkosti lze říci, že známe dvě výběrová šetření:

- **Vyčerpávající šetření** – prošetření všech zkoumaných jednotek/respondentů. V humanitních vědách je toto šetření prakticky nemožné.
- **Výběrové šetření** – v tomto případě pracujeme pouze s výběrem a následně se snažíme generalizovat⁷ na vlastnosti celé populace

Vlastní výběr je možné rozdělit na pravděpodobnostní a záměrný.

7.3.1 Pravděpodobnostní výběry (Disman, 2002)

náhodný – každý respondent může být vybrán se stejnou pravděpodobností (v tomto případě hovoříme o *prostém náhodném výběru*). Dochází tak k tomu, že do vzorku jsou vybráni jak zástupci již známých charakteristik populace, tak zástupci neznámých charakteristik. Jednoduše lze říci, že respondenty náhodně losujeme. Nejlepší způsob, jak provést náhodný výběr je losováním nebo pomocí tabulky náhodných čísel.

⁴ V tomto případě hovoříme o tzv. základním souboru. Můžeme říci, že základní soubor je souhrn prvků, jejichž vlastnosti chceme poznat. Nebo jednodušeji, základní soubor je soubor jednotek, o kterém předpokládáme, že jsou pro něj naše závěry platné. Například se může jednat o všechny žáky základních škol mající trojku z matematiky.

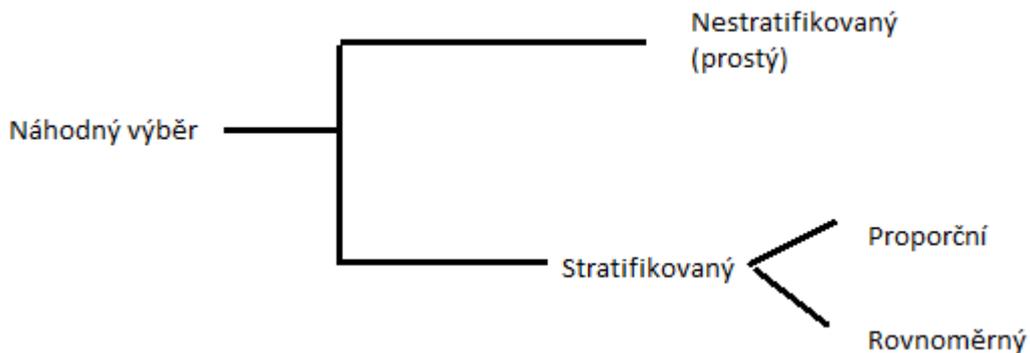
⁵ Jedná se o skupinu těch jednotek/respondentů, s kterou opravdu pracujeme/pozorujeme je.

⁶ Poznámka: Je možné uvažovat jak pravděpodobnostní (charakteristický pro kvantitativní výzkum), tak nepravděpodobnostní výběr (charakteristický pro kvalitativní výzkum).

⁷ Statistika používá postupy, pomocí nichž můžeme, za předpokladu s určitého, předem stanoveného rizika (hladina významnosti), usuzovat na chování celku (populace). Tomuto zobecňování říkáme *statistická indukce*.

Speciálním případem je vícestupňový náhodný výběr. Např. Náhodně vyberu město, v něm náhodně vyberu školu, v ní náhodně vyberu třídu a v této třídě náhodně vyberu žáka.

Dalším speciálním případem je stratifikovaný náhodný výběr. V tomto případě již předem vycházím z nějakého rozdělení populace (muži, ženy, mladí, staří apod.) Tyto skupiny bývají označovány za straty. Jednou takovou skupinou mohou být například adolescenti. Stratifikovaný výběr může vypadat tedy například tak, že z celé populace nejdříve vyselektuji adolescenty a mezi nimi následně náhodně vybírám. (Olecká, Ivanová, 2010). Ukázky způsobu náhodného výběru jsou zobrazeny na následujícím obrázku:



V případě proporčního výběru (někdy také označovaného za kvótní) vycházíme z předpokladu, že již známe nějaké vlastnosti populace (základního souboru). Víme-li například, že na základních školách je např. poměr chlapců a dívek 2:1, musíme vzorek sestavit tak, aby měl stejné charakteristiky. Jinými slovy, i v tomto vzorku musí být poměr chlapců a dívek také 2:1. V rámci kvótního výběru je nutné řešit také počet kvót, druhy kvót, specifické efekty (skupinový, stratifikovaný) a princip školení tazatelů.

- Počet kvót: zpravidla 4-5 kategorií (věk, pohlaví, ekonomická aktivita, lokalita..)
- Druhy kvót: jednoduché (6 žen, 4 muži; 2 studenti, 2 důchodci,..)
složené: propojování kvót (1 žena vysokoškolačka na mateřské, 1 vyučený dělník - zaměstnaný, 1 vyučený dělník v důchodu,...)
- Specifický efekt: **skupinový**: tazatel není 100% spolehlivý při aplikaci náhodného výběru; homogenizace respondentů dle vlastního neuvědomělého úsudku; tazatel si může vybrat skupinku ze svého sociálního okolí

Stratifikovaný: ve výběrovém vzorku mohou být výrazněji odděleny vrstvy, není mezi nimi plynulý přechod; zčišťuje kategorie a posiluje jejich středy (kvóta 25-35 let, tak nejvíc jich tazatel vybere kolem 30ti let); z hlediska reprezentativity to není dobré, ale z hlediska analýzy dat to lépe třídí a lépe to vychází.

Při rovnoměrném výběru jsou osoby v jednotlivých oblastech zastoupeny rovnoměrně. Například stejný počet mužů i žen, i když víme, že v populaci tomu tak není.

- a) **Systematický/mechanický** - První jedinec musí být vybrán náhodně a pak bereme ze seznamu třeba každého 6. jedince. Zjednodušeně lze říci, že je vybrána každá n-tá jednotka ze seznamu.

7.3.2 Záměrný výběr (Hendl, 2005, 2006)

Záměrných výběrů najdeme hned několik. Pro úplnost zmíníme a v krátkosti popíšeme každý z nich. Vzhledem ke skutečnosti, že v praxi se nejvíce využívá kvótní výběr, budeme u jeho popisu značně detailnější. Zdroj ([zde](#))

Kvótní Výběr

Při kvótním výběru **Princip školení tazatele**: rozpis kvót podle schopností a lokality tazatele, tvrdé instrukce, přidělit rajón, který by neměl porušovat, neoslovovat známé; neopakovat se stejnými lidmi; ale zase sdílnější jsou ti respondenti s pozitivním názorem – zakresluje

Výhody: Rychlý, Pružný, Anonymní dotazování, Menší náklady

Nevýhody: Velký vliv tazatele, obtížná kontrola jeho práce

Poznámka: vychází z teze, že jde o kvazireprezentativní výběr (klasické statistické procedury jsou určeny pro náhodný výběr) námitka: pokud se umíme vyhnout chybám, můžeme výběr za reprezentativní označit.

Účelový výběr

Vychází ze záměru výzkumníka. Při tomto výběru musí výzkumník dobře zdůvodnit, proč byl vybrán právě daný vzorek.

Výběr na základě dobrovolnosti

Tato technika je využitelná v podstatě jen v situaci, kdy pracujeme s malým vzorkem respondentů. Výzkumník například sestaví dotazník, vyvěsí jej na web a čeká, až nasbírá dostatečný počet respondentů. V tomto případě však může dojít k problému, že například všichni respondenti jsou z jedné skupiny, všichni jsou muži apod. Základní myšlenkou je, že respondenty necháme dobrovolně reagovat na nabídku, neoslovujeme je přímo a nemanipulujeme jimi.

Výběr na základě dostupnosti

jedná se o výběr, který je nejvíce využíván v bakalářských, diplomových a někdy také v disertačních pracích. Tento výběr vychází z faktu, že ne každý z respondentů je pro studenta dostupný (vzdálenost, neochota apod.) a proto jsou zvoleni ti respondenti, kteří jsou v aktuální chvíli dostupní.

Technika sněžové koule

Jak již název napovídá, tato technika spočívá v tom, že nové respondenty nabíráme na základě doporučení respondentů, kteří již byli testováni.

Poznámka: Mezi další druhy výběrových šetření je možné zařadit anketu nebo metodu základního masivu.

Splnění pravděpodobnostního potažmo záměrného výběru ještě nezaručuje, že vzorek námi vybraný je v pořádku. Stále možné vzorek sestavit nesprávným způsobem a vystavit se tak riziku, že data, s kterými chceme nadále pracovat, budou zkreslená. K tomuto zkreslení může dojít zejména z následujících důvodů:

- a) Malý výzkumný soubor (nereprezentativní) případně jeho špatná korespondence se základním souborem. Ukázka chyby je například výběr lidí podle telefonního seznamu dopředu vyřazuje osoby, které telefon nemají.

Reprezentativitu vzorku shrnul do přehledné tabulky Katriak (1975)

Příklad velikosti základního souboru	Přibližná velikost výběrového souboru
Do 20 respondentů	100 %
Do 100 respondentů	80 %
Do 1 000 respondentů	40 %
Do 10 000 respondentů	7,5 %
Do 100 000 respondentů	1,5 %
Do 1 000 000 respondentů	0,25 %
Do 10 000 000 respondentů	0,06 %

- b) Špatné složení experimentální a kontrolní skupiny (selektivní zkreslení). Toto nastane za předpokladu, že dvě skupiny se neliší pouze v tom, že na jednu je cíleně působeno nebo se u nich sleduje konkrétní faktor, ale také v situaci, kdy se skupiny liší i v jiných faktorech (třeba zaměstnání, věku, početnosti rodiny, životního stylu aj.).
- c) **Nedostatečná velikost zkoumaného souboru** – nejčastější problematika v závěrečných pracích studentů. Vzorek není dostatečně velký, aby pokryl všechna specifika zkoumané populace.

7.4 Tvorba dotazníku/testů

Na tomto místě je nutné zmínit, že je veliký rozdíl mezi dotazníkem a testem. Zatímco dotazník má za úkol zjistit názory respondenta, jeho postoje apod., test jej má analyzovat a tak, jak název říká, podrobit testování.

Poznámka: Student by si měl uvědomit, že zpravidla není odborníkem na danou oblast. Je proto mnohem jednodušší využít již hotový dotazník/test (případně jej přeložit) a ten následně použít. Tyto dotazníky jsou zpravidla již doplněny o manuál k jejich zpracování apod.

Tvorbu správného dotazníku/testu je možné označit za náročnou problematiku, kterou si budoucí „výzkumníci“ potažmo učitelé neuvědomují. Zmiňme například zásady spojené s touto problematikou a ihned zjistíme, že se nejedná o triviální záležitost.

7.4.1 Zásady tvorby dotazníku:

- Otázky jednoduché a srozumitelné (snaha o vyloučení nejednoznačnosti)
- Vyvarovat se obsahové nesrozumitelnosti
- Dotazy musí být co nejvíce specifické
- Minimalizovat dlouhé dotazy
- vyvarovat se příliš strohých otázek
- Nesmí se zde objevovat předsudky
- Kvalitní grafická stránka
- Podrobný návod k vyplnění
- Termín návratnosti
- Nabízené možnosti výběru odpovědí by měly být vyčerpávající a pokrývat všechny možné možnosti.
- Věcná náročnost: respondent nemusí znát odpověď. např. Vyjmenujte všechny příbuzné za pět generací.
- Pokud jsou anonymní, neměly by obsahovat údaje, které by umožňovaly odhalení anonymity.
- Neptat se na více věcí jednou otázkou (Proč podle vás děti nechodí do lesa a proč se zhoršuje jejich školní prospěch?)
- Nepoužívat otázky, na které po pravdě nikdo neodpoví (Čtete? Fetujete? Podvádíte partnera/partnerku?)
- Vyvarovat se sugescí (Také si myslíte, že Tesla není ekologický vůz?)
- Otázky typu Proč? Pokud je možné odpovědět v krátkosti, mohou být tyto položky použity (Proč se věnujete sportu, který jste uvedl/a). Za nesmyslné jsou pak považovány otázky, kde odpovědět je velice zdlouhavá (Jaké je podle vás řešení migrační krize?)
- Otázky spojené s haló⁸ efektem. Zde je problém, pokud klademe příliš podobných otázek a první je spojená s výrazným citovým nábojem potažmo negativní odpovědí. tento postoj se pak přenáší také do dalších položek.

⁸ Za Haló efekt je považováno, jestliže nás jedna nápadná vlastnost člověka ovlivní natolik, že v jejím světle chybně hodnotíme jeho ostatní vlastnosti

Rozmístění otázek

Platí důležitá poučka „**Nejdůležitější je vždy uprostřed**“.



Platí, že otázky obecného rázu mají být na začátku dotazníku a abstraktní umístěny před konkrétními. Sled otázek postupuje od snadných, objektivních a nekontroverzních k složitějším, subjektivním a kontroverznějším.

7.4.2 Typy položek v dotazníku

Při sestavování dotazníku je možné využít celou řadu otázek, jak nadále uvádíme. Ve chvíli, ve které se s těmito typy položek budete seznamovat, mějte na mysli, že daný dotazník chcete umět také adekvátně vyhodnotit. Pokud neholdujete kódování (například dle Švaříčka a Šed'ové) případně nemáte k dispozici Alas.cz, pak budete otevřené otázky vyhodnocovat obtížně.

Uzavřené

Výhodou uzavřených otázek je, že máte již předem formulované odpovědi a otázka se tak snadno zpracovává. Práce je to pro vás sice jednoduchá, ale mějte na mysli, že zde hrozí nebezpečí zkreslení. Tyto otázky je tedy vhodné doplnit o možnost jiné odpovědi. Výhody a nevýhody uzavřených otázek je možné přehledně shrnout do následující tabulky:

Snadnost a rychlost vyplnění	Náročná formulace otázky a možností
Snadné zpracování a vyhodnocení odpovědí	Není vhodná pro otázky s velkým množstvím odpovědí
Vtažení respondenta do problému otázky	Nucené odpovědi respondentů (nemožnost vlastní volby respondenta)
Velká pravděpodobnost vyplnění	

Nadále uvádíme popis jednotlivých typů otázek a u každé z nich uvádíme příklad. Tento příklad je prostou kopií obrazovky ze známého serveru survio.com.

a) Dichotomické (s alternativou)

Zde je možné si vybrat pouze ze dvou navzájem se vylučujících odpovědí. Tyto typy otázek jsou velice omezené a nedoporučuje se sestavit celý dotazník pouze z tohoto typu otázek.

Ukázka:

Jste ve svém zaměstnání spokojen/a?

Ano

ne

Někdy se setkáme také s pojmem trichotomická otázka, která bývá zařazena mezi polytomické.

Ukázka:

Jste ve svém zaměstnání spokojen/a?

Ano

Nevím

Ne

b) Polytomické výběrové

V tomto případě je předloženo vícero možných odpovědí s tím, že pouze jedna je správně. Je nutné, aby volba jedné z odpovědí automaticky vylučovala jinou odpověď.

Ukázka:

Jakou barvu by podle Vás měla mít moderní lednice?

Bílou

Černou

Červenou

Hnědou

Žlutou

Ani jednu z uvedených

c) Polytomické výčtové

Tyto otázky jsou obdobné, jako (b) s tím rozdílem, že je možné vybrat více odpovědí. Zde je důležité, aby respondent věděl, že je možné vybrat více odpovědí a také, že není určující pořadí těchto odpovědí.

Ukázka:

Jak hodnotíte našeho obchodního zástupce?

Příjemný

Vstřícný

Sympatický

Zkušený

Upravený

Důvěryhodný

d) Polytomické stupnicové

Jako (c) s tím rozdílem, že pořadí odpovědí je již určující. Klasickým případem je seřazení školních předmětů podle oblíbenosti.

e) Stupnicové komparativní

Jedná se o velmi oblíbené typy otázek, v rámci kterých respondent udělí své odpovědi hodnotu podle uvedených možností. Velice oblíbené je Likertovo⁹ škálování znázorněné na následujícím obrázku:

Jak se vidíš?

A	Odvážný	1	2	3	4	5	6	7	Ustrašený
B	Starostlivý	1	2	3	4	5	6	7	Lhostejný
C	Vyhovavý	1	2	3	4	5	6	7	Nevychovavý
d	zručný	1	2	3	4	5	6	7	nešikovný

Dogma: Obecně platí, že tato škála by měla být polarizována pomocí dvou antagonistických pojmů a obsahovat lichý počet stupňů.

V praxi se setkáte s názory, že je nutné mít indiferenční bod (střed) a také s názory, že není nutné jej mít¹⁰. Podívejme se na dva názory týkající se indiferenčního bodu. Například (Pacce-Bucci, 2013) tvrdí, že absence bodu indiference nemá dopad na výsledky průzkumu, i když připustíme, že u určitých prvků se bude muset respondent přiklonit na jednu nebo druhou stranu, ačkoliv o tom nebude subjektivně přesvědčen. Naproti tomu Dumas tvrdí, že spolehlivost škály je absencí neutrálního bodu zcela zásadně narušena, čímž je snížena i věrohodnost dosažených výsledků.

Za větší problém považujeme nabídku možných odpovědí, jelikož ta nám, do jisté míry, předurčuje nebo podsouvá naši odpověď. Než tedy psát běžné (Souhlasím, spíše souhlasím, nemám vyhraněný názor, spíše nesouhlasím, nesouhlasím), je vhodné využít stupnici nebo se držet moderního trendu a tedy použít úsečku, na kterou bodem respondent nanese míru souhlasu s danými antagonistickými tvrzeními.

Otevřené

Tyto položky mají nespornou výhodu v tom, že je na jejich základě možné poměrně přesně postihnout názor respondenta, jelikož se jedná o položky s volnou odpovědí. Problematika spočívá v jejich zpracování. Představte si, že posbíráte data od 200 respondentů, kdy u každého z dotazníků bude jedna položka otevřená a v rozsahu poloviny strany. V tuto chvíli máte 100 stran textu k „okódování“. Pokud si nejste jisti, že se v dané oblasti orientujete, vůbec se do těchto položek nepouštějte.

⁹ Likertovo škálování je metodou, která je používána pro určení míry stupně souhlasu či nesouhlasu s tvrzením, se kterým jsou respondenti výzkumu konfrontováni. Likertova škála, vyvinuta roku 1932, představuje jednu z nejspolehlivějších technik měření postojů (Hayes, 2003).

¹⁰ Často se také lze setkat se sémantickým diferencialem, kdy respondent hodnotí na obvykle sedmibodové stupnici například určitou vlastnost za využití dvou antonym, přičemž uprostřed je neutrální postoj a krajní hodnoty se vždy s určitou mírou přiklánějí k jednomu z opozit (Pecáková 2008, s. 36).

Poznámka: Tento typ odpovědí se používá spíše u kvalitativního výzkumu, kde je podstatně méně respondentů.

Polouzavřené

Jedná se o další typ oblíbených otázek, které jsou vystavěny na tom principu, že pokud respondentovi nevyhovuje žádná z nabízených odpovědí, použije vlastní možnost. Polouzavřená otázka vznikne přidáním varianty "jiné" do uzavřené otázky, která je vlastně otevřenou otázkou a umožňuje respondentovi volně vyjádřit svůj názor.

Příklad: Z jaké země pocházíte?

- Česká republika
- Slovenská republika
- Polsko
- Maďarsko
- jiné:

Ukázka:

Zdroj: <http://www.dotaznik-online.cz/otazky-dotazniku.htm>

Filtrační

Tyto položky se často používají pro ověření určité odpovědi. Klasická ukázka je zde: „Pokud jste odpověděl ano, pak uveďte.....“. Právě odpověď ano-ne bývá často použita jako filtr, ale nemusí se jednat o dogma. Ukázka: „Jste sportovec ANO-NE? Pokud ano, jakému se věnujete sportu?“

Projekční

Tyto položky vycházejí, jak již název praví, z projekce¹¹ a velice často se používají v sociální psychologii. Vychází se například z předpokladu, že než říci, že kradu je jednodušší říci, že každý někdy něco ukradl. Lidé často vlastní názory projektují na ostatní.

Ukázka (dotaz na učitele):

Jak myslíte, že jsou učitelé spokojeni se svými platy?

Ukázka (nedokončené věty):

Petr okradl Jitku, když stála ve frontě. Jitka.....

¹¹ Někdy je možné setkat se s označením projektivní místo projekční.

Kontrolní

Používají se zpravidla k ověření odpovědí respondentů a jsou umístěny daleko od otázky, kterou mají ověřit.

Ukázka: Kouříte (ANO – NE) Jaké kouříte cigarety?

7.4.3 Závěrem k tvorbě dotazníku

Co musí každý dotazník splňovat?

Co musí stát v úvodu dotazníku:

Oslovení respondenta, účel výzkumu, představení tazatele (jméno, tituly, škola, studium), anonymita – většinou u jednorázových dotazníků, instrukce z vyplnění, poděkování a podpis.

Poznámka: Nezapomeňte, že daný dotazník budete také zpracovávat a je tak nutné mít na mysli determinující proměnné jako je věk, pohlaví, vzdělání apod.

V rámci dotazníku je možné použít celou řadu otázek. Nadále uvádíme jejich výčet.

Výhody:

- velké množství dat získatelné v krátkém čase
- nedochází k ovlivňování
- elektronická forma je ihned zpracovaná

Nevýhody:

- kontrola
- návratnost
- téměř žádná spontaneita
- dotazník dává často obrázek o tom, co respondent říká a jaký chce být, než o tom, jak myslí a jaký opravdu je. Giddnes (2005)

7.4.4 Validita dotazníku

Validní měření = platné měření tj. takové měření, které měří skutečně to, co jsme zamýšleli měřit. Zabývat se validitou znamená podle Gavory (2000) řešit otázku, zda má daný nástroj (evaluační nebo výzkumný) schopnost zjišťovat to, co zjišťovat má.

Jedná se o míru, do jaké test měří skutečně to, co má. Zdrojů důkazů o validitě daného konstruovaného nástroje je celá řada. V této kapitole budou zmíněny nejzákladnější z nich. Podrobně se této problematice věnuje Schindler (2006). O validitě lze tedy zjednodušeně říci následující:

Validita představuje shodu mezi výsledky testu a účelem, pro který byl test vytvářen (Schindler, 2006). Standardy pro pedagogické a psychologické testování (Praha, Testcentrum, 2001) definují validitu jako míru, ve které empirické důkazy a teorie podporují interpretaci testových skóre při doporučeném způsobu užívání testu.

Poznámka: Validitu je možné považovat za nejdůležitější údaj přiřazený k danému testovanému nástroji. Pokud není validita splněna, není možné nástroj použít a počítat tak jiné údaje jako reliabilitu, objektivitu¹² apod. je zbytečné.

v případě testování validity se autoři mnohdy liší. Ne snad v tom, jak se testování provádí, ale spíše v jeho dělení. V této kapitole se zaměříme na tři základní způsoby validizace nástroje a těmi jsou obsahová, konstruktová a kriteriální validizace.

Obsahová validita

Obsahová validita je zaměřena speciálně na vědomostní testy a udává, do jaké míry jsou dané testové položky vhodným reprezentativním zastoupením daného problému. Zaměřuje se tak na úplnost významové domény daného jevu a zjišťuje, do jaké míry měření skutečně reprezentuje dané vlastnosti a kvality. Obsahovou validitu je možné zkoumat například odkazem na relevantní literaturu, jiným empirickým výzkumem stejného charakteru nebo na základě mínění expertů z dané oblasti. Zjednodušeně je možné říci, že obsahová validita zjišťuje, zda jsou položky testu opravdu zaměřené pouze na danou problematiku a pro jejich vyřešení již není nutné dalších vědomostí. Demonstrujme tuto problematiku na jednoduchém příkladu společně s komentářem, který uvádí Schindler a kol. (2006).

Obec Kovač má 250 obyvatel. Každý rok se v obci narodí nebo se do ní přistěhují 3 obyvatelé. Každý rok 2 obyvatelé zemřou nebo se odstěhují. Vypočítejte, za kolik let se obec Kovač stane městem.

KOMENTÁŘ: Cílem úlohy je ověřit základní matematické operace. Při řešení úlohy žáci mají uvést postup výpočtu a závěr, který obsahuje počet let, za něž se daná obec stane městem. V zadání úlohy však není uvedena definice města. Předpokládá se, že žáci vědí, jaké podmínky musí být splněny, aby se obec změnila v město. Tato znalost však jistě nepatří do matematiky. Úspěšnost žáků při řešení úlohy tedy nezávisí pouze na znalostech a dovednostech matematických.

¹² Objektivita je pojem, který souvisí s didaktickými testy, které by měly být vyhodnocovány co možná nejobjektivněji. Současným trendem je ovšem odstup od položek čistě objektivně hodnotitelných. Přesto platí, že test by měl být takový, aby se shodovala hodnocení od různých posuzovatelů (učitelů) na základě daných kritérií.

Zdrojů nízké obsahové validity testu je hned několik:

1. Pro vyřešení úlohy je nutné mít informace, které doposud nebyly probrány.
2. Pro vyřešení úlohy je nutné mít informace, které nesouvisí s danou problematikou.
3. Formulace dané položky.
4. Nejednoznačnost vyjadřování.
5. Omezení žáka (SPU, handicapovaný aj.)

Zjištění obsahové validity testu:

V praxi se obsahová validita testu zjišťuje pomocí odborníka nebo skupiny odborníků. Tato validita bývá někdy nazývána zjevná (face validity), jelikož se zjišťuje na základě prostého úsudku expertů o validitě měření. K tomuto se vztahuje základní chyba některých bakalářských potažmo diplomových prací studenta. Žák zpravidla sestaví test tak, aby byl schopen jej dobře zpracovat případně rozeslat v elektronické formě. Tento test prodiskutuje s vedoucím, který mu jej schválí a spokojený student se pouští do práce. Nesmíme však opomenout, že v případě obsahové validity testu se očekává hodnocení (nejlépe písemné) od odborníka¹³ z dané oblasti. Skutečnost, že má vyučující docenturu z matematiky, ještě neznamená, že je schopný objektivně posoudit test zaměřený na motivaci. S velkou nadsázkou řečeno: Gynekolog také objektivně neposoudí test zaměřený na rozvoj prostorové představivosti.

Důležité na závěr:

- Obsahová validita se stanovuje posudkem experta/expertů z dané oblasti.
- Validita není něco pevného, ale mění se v závislosti na podmínkách zjišťování (výzkumu). Tj. pokud někdo validizoval nástroj pro měření prostorové představivosti studentů střední škol, není možné ten samý nástroj použít pro žáky základních škol a považovat jej za dostatečně validní!
- Za validitu je podle novějšího pojetí (Chvál, 2003) zodpovědný nejen autor daného nástroje, ale i ten, kdo ho používá a kdo interpretuje výsledky konkrétního zjišťování.

Konstruktová validita

Konstruktová validita – vyjadřuje míru vztahu mezi testem a konceptem (teoretickým konstruktem), který bývá zpravidla modelován pomocí latentní proměnné. Číselné vyjádření není tedy přesné a můžeme ji pouze odhadovat například pomocí faktorové validity, kdy jeden z modelů konstruktové validity, vyjadřuje validitu testu k latentnímu kritériu, které se uměle konstruuje z testových výsledků baterie testů pomocí modelu faktorové analýzy

Konstruktová validita je doložena, pokud je v datech zjištěn takový vztah mezi daným indikátorem a dalšími proměnnými, jaký bychom a priori očekávali na základě teorie. Pokud

¹³ Definovat odborníka v dané oblasti je sporné. Zpravidla je za něj považován člověk, který publikuje na dané téma v posledních letech v uznávaných periodikách.

jsou v datech zjištěny vztahy odpovídající teorii, je toto zjištění bráno jako potvrzení konstruktové validity zkoumané proměnné [např. Bagozzi 1978; Edmundson, Koch 1993; Jöreskog, Sörbom 1975; Ruble, Stout 1990; Williams 1994].

Základním nástrojem k měření konvergentní a diskriminační validity jsou explorační a konfirmační faktorová analýza a jejich různé aspekty. Posledním, a teoreticky i etodologicky nejnáročnějším, přístupem k měření konstruktové validity je potom tzv. MIMIC model, který umísťuje zkoumaný měřicí nástroj, zpravidla škálu, do kontextu jak konceptuálně příbuzných, tak odlišných proměnných, přesně specifikuje možné kauzální vazby mezi proměnnými a kontrastuje v datech zjištěný vzorec vztahů s teorií. Pokud jsou v datech zjištěny vztahy odpovídající teorii, je toto zjištění bráno jako potvrzení konstruktové validity zkoumané proměnné.

Kriteriální validita

Jedná se o validitu, která je pro studenty nejsnáze počitatelná (pokud se nespokojí s názorem vedoucího, jež považují za experta – obsahová validita). Kriteriální validizace je založena na shodě **mezi výsledky výzkumného nástroje a výsledky jiného měření. Tato validita se odhaduje zkoumáním vztahu výsledku testu k danému nezávislému kritériu. Jak lze tušit, často největším problémem prokazování kriteriální validity je právě nalezení vhodného kritéria.**

Existuje hned několik přístupů ke kriteriální validizaci:

Kriteriální validita souběžná (Hendl, 2012)

Souběžná kriteriální validita nástrojů se zjišťuje jejich vzájemnou korelací. V praxi probíhá testování v několika krocích:

1. Student sestaví/přeloží/převzme výzkumný nástroj.
2. student sežene jiný výzkumný nástroj zaměřený na tutéž problematiku jako jím sestavený/přeložený převzatý, který je již validizovaný.
3. Student rozdá oba nástroje stejné skupině respondentů a zjišťuje míru shodu získaných dat. Pokud nově sestavený/přeložený/převzatý nástroj silně koreluje s nástrojem, o kterém víme, že je již validizovaný, je možné očekávat, že i tento nový nástroj bude také validní.

Poznámka: Validizaci je nutné chápat jako proces vedoucí k „validitě“; jakmile tedy něco úspěšně validizují, považují to za validní.

Kriteriální validita predikční

„Predikční kriteriální validita je založena na představě o tom, do jaké míry náš test **předpovídá budoucí hodnoty nějakého kritéria.** Predikční validita vypovídá o tom, do jaké

míry dokáže náš test předpovídat výsledky v budoucnu v oblastech, které nás primárně zajímají a kvůli kterým je test konstruován (např. souvislost mezi výsledkem testu studijních dovedností pro 9. ročníky ZŠ a úspěšností v dalším studiu.). (Schindler, 2006)

Validita inkrementální (přírůstková)

„Jedná se o validitu zjišťující, zdali daný test **přináší novou informaci nad tu, kterou získáme jiným prováděným testem.** V případě zmíněných přijímacích testů nás může například zajímat, zda přijímací testy přidávají novou informaci o budoucím studiu uchazeče nad tu, kterou nám poskytne jeho středoškolský prospěch. Např. studie (Štuka, 2012) na základě dat studentů přijatých na 1. LF UK ukázala, že středoškolský prospěch vysvětlí zhruba 15 % variability úspěšnosti ve studiu. Výsledek z přijímací zkoušky zvýší procento vysvětlené variability úspěšnosti na 22 %, přidání informace o úspěšně absolvovaných profilových předmětech na střední škole na 25 % a informace o roku maturity (v roce přijetí nebo dříve) dokonce na 30 %. Všechny zmíněné efekty byly v modelu signifikantní (tedy statisticky průkazné), prokázala se tak jejich přírůstková validita“. (Schindler, 2006)

Faktory ovlivňující validitu

Reliabilita určuje spolehlivost daného testování a bývá charakterizována následovně.

Reliabilita je komplikovaná veličina zahrnující v sobě **spolehlivost a přesnost.**

Spolehlivost je možné popsat tak, že pokud daný test zadám dvěma skupinám respondentů totožných schopností/dovedností a za totožných podmínek, dostanu také totožné výsledky. Spolehlivé měření znamená, že test poskytuje stabilní, opakovatelné výsledky. Pokud test neměří spolehlivě, znamená to, že do výsledků se promítá řada vnějších, náhodných vlivů.

Naopak přesnost bývá vnímána tak jako vzdálenost mezi skutečnými znalostmi a dovednostmi žáka a jeho dosaženými výsledky v testu, která se může měnit. Jestliže je test dostatečně přesný, je tato vzdálenost malá a výsledek v testu je pak téměř totožný se žákovými skutečnými znalostmi a dovednostmi.

Na základě výsledků málo reliabilních testů není možné činit významné závěry o žácích, protože nelze vyloučit příliš velký podíl náhody. (Schinder, 2006).

Reliabilitu je možné v zásadě spočítat hned několika způsoby:

1. Test-retest reliabilita
2. Reliabilita paralelních měření
3. Reliabilita testu jako vnitřní konzistence
 - i. Splithalf reliabilita – korekce pomocí SpearmanBrownova vzorce
 - ii. Cronbachova alfa – ordinální položky
 - iii. Vzorec Kudery a Rychardsona č. 21 nebo 20 – dichotomické položky

Nadále se věnujeme podrobnějšímu popisu každého ze způsobů ověření reliability výzkumného nástroje:

Test retest reliability

Jak již název napovídá, tento typ reliability říká, jak moc spolu koreluje ten samý test při opakovaném použití na těch samých lidech (měli bychom vždy naměřit to samé). Zpravidla se využívá běžných korelací. Tento typ reliability má dvě zjevné nevýhody:

- Při administraci po uplynutí příliš krátké doby vzniká tzv. efekt zácviku a nadhodnocení reliability kvůli zapamatování si položek. Doporučuje se proto rozestup minimálně 3 měsíce.
- Většina psychických vlastností je v čase proměnlivá, proto s větším časovým rozestupem se nutně musí lišit i dosažený výsledek, což reliabilitu snižuje.

Jak tedy provést měření na základě test-retest reliability:

1. Změřím danou skupinu respondentům pomocí nástroje, který ověřuji (zpravidla dotazník).
2. Sleduji skupinu po dobu několika měsíců (minimálně tři) a to z toho důvodu, abych věděl, že do hry nevstupuje nežádoucí faktor¹⁴.
3. Provedu opětovné šetření za využití stejného nástroje s tím, že nezapomenu okomentovat, zda je možné, aby došlo ke změně psychických vlastností v čase (vývojem jedince).

Testy, které měří poměrně stabilní psychické vlastnosti (inteligence, extraverte apod.) se mohou problémům spojeným s vývojem jedince vyhnout.

Reliabilita paralelních měření

Tato reliabilita vychází z předpokladu, že výzkumník vytvoří dvě verze testu měřící stejnou psychickou vlastnost, které administruje jediné skupině respondentů a vzájemně je pak koreluje. Předpokládá se také, že oba testy mají stejné chybové rozptyly. Následně pak z pohledu teorie testování, že korelace mezi dvěma paralelními testy

$$X_1 = T + E_1$$

$$X_2 = T + E_2$$

je rovná jejich reliabilitě.

V praxi bývá problém s tím, že se obě formy testů liší obsahem, případně, že neměří totožnou psychometrickou vlastnost.

¹⁴ Budu-li například testovat prostorovou představivost a ta bude tři měsíce cíleně rozvíjena, lze očekávat, že dojde k posunu a tedy k odlišným výsledkům testu.

Reliabilita testu jako vnitřní konzistence

Základní myšlenka testování vnitřní konzistence testu je jednoduchá. Předpokládá se, že různé části celku (jednotlivé položky¹⁵ případně skupiny položek) je možné zkoumat jako různé formy testu. U vnitřní konzistence předpokládáme, že každá úloha měří pouze tu znalost či dovednost jako celý test. Pro vnitřní konzistenci platí, že čím více úloh test obsahuje a čím jsou si úlohy podobnější, bližší (tj. ověřují co nejužší skupinu znalostí a dovedností), tím vyšší má reliabilitu. Platí fakt, že jakmile tyto položky mezi sebou silně korelují (jsou silně spjaté) je pravděpodobné, že testovaná skóre jsou konzistentní se skórem testu jako celku a ten je pak možné pokládat za reliabilní. Platí, že delší test vykazuje větší reliabilitu, než test o několika otázkách (této problematice se blíže věnujeme v kapitole Báje a mýty o reliabilitě).

Poznámka: Jak zmiňuje Schindler (2006) „*Snaha o dosažení vnitřní konzistence jde však proti obsahové validitě testu, kdy se snažíme naopak pokrýt úlohami širší obsahovou oblast a předem víme, že každá úloha měří něco trochu jiného, a na výsledek testu nahlížíme jako na kombinaci požadovaných znalostí a dovedností. V takovém případě případně by bylo vhodnější volit jinou cestu odhadu reliability, např. testování pomocí paralelních forem testů (Říčan, 1978)*“.

Nadále uvádíme tři způsoby odhadu/výpočtu vnitřní konzistence testu:

Splithalf reliability

Tento způsob ověření reliability poprvé použili Johnson & Penny (2005). V průběhu výpočtu se test rozděluje na dvě přibližně stejné části a zjišťuje se jejich korelace. Nevýhoda podstatného snížení hodnoty koeficientu korelace u Split-half reliability se koriguje použitím Spearman-Brownova vzorce, který stanovuje reliabilitu pro celý nezkrácený test. (Johnson & Penny, 2005). Tento vzorec je ve tvaru:

$$r_{x,x'} = \frac{m \cdot r'_{x,x'}}{1 + (m-1) \cdot r'_{x,x'}} \quad (1)$$

kde m je poměr počtu otázek původního testu k počtu otázek testu zkráceného a r je korelace.

Největší nevýhodou metody půlení je závislost koeficientu reliability na tom, jakým způsobem test rozpůlíme. Vhodné je tedy udělat všechna možná půlení, kterých však v praxi může být velké množství.

¹⁵ Pokud hovoříme o položkách, máme na mysli testové otázky.

Cronbachova alfa

Koeficient Cronbach α (Cronbach, 1951; McGartland Rubio, 2005), zjišťuje vnitřní konzistenci nástroje a může nabývat hodnot v intervalu $\langle 0,1 \rangle$, přičemž obecně akceptovatelné hodnoty koeficientu jsou mezi 0,7 a 0,95 (Tavakol & Dennick, 2011). Cronbachovo alfa je populárním odhadem a funkce pro jeho výpočet je obsažena ve většině statistických balíků (SPSS, Statistica v12 apod.).

Pro možnost výpočtu Cronbachovy alfy je nutné znát informace o každé položce zvlášť. Není tak vhodné například využít elektronické formy dotazníků jako jsou survio apod., které automaticky testy vyhodnocují, ale nenabízí matici dat. Představme si, například, že máme test o deseti položkách, kde na každou z nich se odpovídá v intervalu 1-5. Pro výpočet reliability je nám k ničemu, že se dozvíme průměr, medián apod. případně, že budeme mít k dispozici základní grafy, jak to na mnoha serverech bývá (ukázka – obr. 1).

Je nutné získat matici dat, jejíž ukázka je na následujícím obrázku.

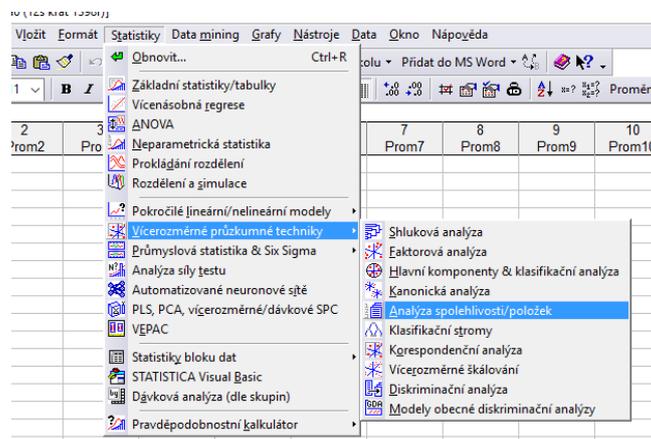
	A	B	C	D	E	F	G	H	I	J	K
		Položka 1	Položka 2	Položka 3	Položka 4	Položka 5	Položka 6	Položka 7	Položka 8	Položka 9	Položka 10
Respondent 1		5	3	5	5	5	1	1	4	4	3
Respondent 2		4	5	4	5	1	4	3	3	2	2
Respondent 3		5	4	5	4	1	4	3	4	2	1
Respondent 4		5	5	3	4	3	1	1	2	4	4
Respondent 5		2	4	2	2	5	1	5	1	4	4
Respondent 6		1	5	1	5	3	5	1	2	2	1
Respondent 7		3	1	5	5	4	2	3	5	2	3
Respondent 8		2	5	3	1	3	4	2	1	2	5
Respondent 9		5	1	5	2	4	2	4	4	1	4
Respondent 10		5	4	3	4	3	4	2	2	4	1
Respondent 11		1	1	2	4	4	3	2	4	4	3
Respondent 12		1	4	3	4	3	1	2	5	3	2
Respondent 13		5	1	1	5	4	3	3	5	4	4
Respondent 14		3	3	5	2	4	5	5	2	2	1
Respondent 15		4	1	4	3	1	1	3	4	1	5
Respondent 16		5	4	3	1	1	3	2	2	4	2
Respondent 17		4	3	3	3	2	1	1	4	5	1
Respondent 18		4	2	4	4	2	1	5	1	1	4

Výpočet Cronbachovy alfy je pak možný na základě následujícího vzorce:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_t^2} \right), \quad (2)$$

kde k je počet položek, σ_i^2 je rozptyl i -té položky a σ_t^2 je rozptyl celého testu.

Výpočet pomocí softwaru Statistica je jednoduchý. Stačí zadat cestu, kterou zobrazuje následující obrázek:



Následně se jako proměnné zvolí všechny položky testu a následně dvakrát zvolí OK.

Rutinní použití nebo vyžadování tohoto způsobu měření reliability však není vhodné, jelikož odporuje doporučením obsaženým v některých publikacích¹⁶, kde se zdůrazňuje, že neexistuje jediný optimální odhad, naopak vědec musí pečlivě zvážit, kterou metodu odhadu reliability zvolí. V následujících sekcích se podíváme blíže na situace, kdy použití Cronbachova alfa vhodné není, a nabídneme alternativní metody odhadu reliabilit.

Více latentních proměnných

Jak bylo zmíněno, Cronbachovo alfa předpokládá, že všechny položky měří jedinou latentní proměnnou. Tento předpoklad lze považovat za splněný např. u úloh testujících algebraické schopnosti, jako je schopnost sčítat a odečítat do 100. Naproti tomu u slovních úloh, které testují jak znalost elementární algebry, tak schopnost číst a orientovat se v textu, je správnější uvažovat dvě latentní proměnné. V psychologii je častý ještě vyšší počet latentních proměnných. Nejsou-li současně porušeny další předpoklady, bude Cronbachovo alfa podhodnocovat skutečnou reliabilitu (Novick, Lewis, 1967).

Korelované chyby

Korelované chyby položek si lze snadno představit v případě, kdy se část otázek vztahuje k témuž uvozujícímu textu. Takové skupiny položek jsou běžné v testech studijní připravenosti nebo např. testech cizího jazyka, jako je TOEFL. V článku Green, Yang (2009) jsou popsány také další možné příčiny narušení předpokladu nekorelovaných chyb, které mohou vést k podhodnocení reliability: např. výrazně omezený čas na vyřešení testu (tzv. rychlostní test), nebo skutečnost, že předcházející položky ovlivňují položky následující.

Případ více zdrojů

Model, který stojí za Cronbachovým alfa také předpokládá pouze jediný typ chyb – náhodnou chybu měření. V praxi je situace často mnohem složitější. Studenty např. může hodnotit více hodnotitelů, studenti mohou test skládat na různých místech s odlišnými podmínkami apod.

Nesplnění normality

Odhad Cronbachova alfa vychází z představy normálního rozdělení jednotlivých měření. Jeho použití tedy může být nevhodné, je-li tento předpoklad porušen. Robustní verzi Cronbachova alfa se zabývá článek Christmann, Van Aelst (2006).

Případ binárních/dichotomických položek

V tuto chvíli se využívá vzorec Kudery a Richardsona popsany v následující kapitole.

¹⁶ Standardy pro pedagogické a psychologické testování. Praha: Testcentrum, 2001

4.5.3.3 Vzorec Kudery a Richardsona č. 21

Tento vzorec se využívá ze předpokladu, že pracujeme s dichotomickými položkami. Je možné využít jak Kuder-Richardsonův vzorec 21, tak Kuder-Richardsonův vzorec 20 – KR-20 (Kuder & Richardson, 1937), jehož hodnoty a výsledná interpretace je obdobná jako v případě koeficientu Cronbach α . Jde o historicky starší koeficient, než Cronbachovo alfa, a ve většině případů poskytuje obdobné výsledky. Jeho výhodou je nižší výpočetní náročnost, díky které lze snadno spočítat i ručně.

Vzorec KR20 je obecně ve tvaru:

$$r_{xx'} = \frac{n}{n-1} \cdot \frac{\sigma_x^2 - \sum_{j=1}^n p_j q_j}{\sigma_x^2}, \quad (3)$$

kde n je počet položek testu, σ_x^2 jejich rozptyl, p_j podíl osob, které mají j -tou položku zodpovězenou správně a q_j podíl osob, které ji správně nezodpověděly¹⁷; $p_j q_j$ je proto rozptyl dichotomické položky.

Vzorec KR21 je obecně ve tvaru:

$$KR_{21} = \frac{K}{K-1} \cdot \left[1 - \frac{AP \cdot (K - AP)}{K \cdot S^2} \right], \quad (4)$$

kde K představuje počet položek dané zkoumané oblasti (například u hledání číselných zákonitostí je $K=6$), AP je aritmetický průměr jednotlivých získaných hodnot v dané oblasti a S^2 je rozptyl daného souboru.

Poznámka: Koeficient KR-21 je specifikací výše uvedeného výpočtu (3) a je určený pro test, ve kterém všechny položky mají přibližně stejnou obtížnost.

7.5 Časté chyby při statistickém zpracování dat

Chyb, kterých se výzkumník v průběhu výpočtu či úvah může dopustit je celá řada. V této části se zaměříme na ty, které jsou buď nejčastější, nebo je jejich výskyt velkým problémem.

7.5.1 Chyba prvního a druhého druhu

Tyto dvě chyby se navzájem doplňují. Chyba 1. druhu: nulová hypotéza sice platí, ale my ji zamítáme. Rozhodnutí o zamítnutí H_0 je dáno hladinou významnosti testu, což je maximální přípustná pravděpodobnost chyby 1. druhu. Hladina testu se zpravidla značí symbolem α . Většinou volíme hladinu významnosti $\alpha=0,05$ nebo $\alpha=0,01$.

¹⁷ Hovoříme-li o vědomostních testech.

K chybě 2. druhu dochází, když nulová hypotéza neplatí, ale my ji nezamítáme (Nepoznáme, že neplatí). Doplněk pravděpodobnosti chyby 2. druhu do jedničky ($1 - \beta$) se nazývá síla testu. Je to pravděpodobnost, že nulovou hypotézu zamítneme, když tato hypotéza neplatí, tedy pravděpodobnost, s jakou neplatnost hypotézy objevíme.

Tyto chyby je možné jednoduše shrnout do následující tabulky:

H_0 je ve skutečnosti \longrightarrow Já se rozhodnu \downarrow	Správná	Nesprávná
Zamítám H_0	Chyba 1. druhu	Správné rozhodnutí
Nezamítám H_0	Správné rozhodnutí	Chyba 2. druhu

Poznámka: Platí, že čím větší nároky kladu na α ($0.05 \rightarrow 0.01 \rightarrow 0.001$), tím vyšší bude β . Platí také, že β klesá také s rostoucím počtem pozorování.

Důležité:

Zjištění statistické významnosti nikdy nemůže nahradit rozhodnutí o vědeckém (věcném) významu výsledků!

7.5.2 Korelace vs. kauzalita

Velmi častým omylem studentů je nepřiměřené využívání statistiky například v oblastech, kde toto předem postrádá význam. Typicky se tak stává, že bývá zaměňována korelace s kauzalitou. Představme si situaci, kterou popisuje Meloun (2011) v učebním textu s názvem "Počítačová analýza vícerozměrných dat v oborech přírodních, technických a společenských věd".

Příklad: "Statistické ročenky o demografii ukazují, že například počet narozených dětí na vesnicích ve Skandinávii koreluje s počtem čápů vyskytujících se v tomto kraji s korelačním koeficientem $r=0,75$ ".

Každý, kdo není náměstkem ve známém filmu Policejní akademie, jistě tuší, že čáp děti nenosí a tak nelze předpokládat, že počet narozených dětí je dán počtem čápů v dané oblasti. Platí, že korelace nemusí být vždy totožná s kauzalitou. Co je ovšem možné říci s jistotou:

- Korelace je nutným předpokladem kauzality

7.5.3 Chyba měření

Při interpretaci dat nesmíme zapomenout na to, jak byla data získána. Pokud budou data získána na základě výběru (výběrová data), je třeba pamatovat na skutečnost, že se jedná pouze o odhady, které jsou zatížené určitou chybou a nejedná se tak o přesná čísla. Tato chyba má složku výběrovou a nevýběrovou:

Výběrová: K výběru respondentů dochází náhodně (bez jejich vracení), vybírá se pouze jeden náhodný, výběrový soubor a údaje z tohoto souboru nerepresentují základní soubor. Tato chyba vzniká v důsledku ztráty informace při výběrovém způsobu zjišťování, kdy populace je zastoupena výběrem

Nevýběrová: Tato chyba se vyskytuje kdekoli v průběhu zpracování dat a to i za předpokladu, že se jedná o vyčerpávající výzkum. Tato chyba se dá ovlivnit precizní prací, spoluprací více výzkumníků ap. Vzniká zejména z následujících důvodů:

- Špatná přípravná práce v rámci výzkumu (konceptualizace, operacionalizace)
- Nepřesné informace ze strany respondentů spojené s neochotou nebo dalšími faktory jako je stud, aktuální stav ap. (snížená validita).
- Špatná metodika práce
- Špatný postup při zpracování dat (t-test není samospásný a není možné jej aplikovat na všechna možná data)

Zjednodušeně je možné říci, že nevýběrová chyba je způsobena porušením "pravidel hry".

Chybějící hodnoty: i když nejde v pravém slova smyslu o chyby v datech, musíme počítat s tím, že někdy se nepodaří dohledat všechny údaje a jejich část chybí. Na tuto skutečnost v praxi reagujeme různými metodami „ošetření“ chybějících hodnot.

Mezi další možné chyby je možné zařadit:

- Chyba metody
- Chyba v pozorování
- Chyby při vyhodnocování