

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

BAKALÁŘSKÁ PRÁCE

Reliabilita a validita - různé pohledy na kvalitu
psychologických testů



Katedra matematické analýzy a aplikací matematiky

Vedoucí bakalářské práce: **RNDr. et PhDr. Ivo Müller Ph.D.**

Vypracoval(a): **Michal Bechný**

Studijní program: B1103 Aplikovaná matematika

Studijní obor: Matematika-ekonomie se zaměřením na bankovníctví/pojišťovnictví

Forma studia: prezenční

Rok odevzdání: 2017

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Michal Bechný

Název práce: Reliabilita a validita - různé pohledy na kvalitu psychologických testů

Typ práce: Bakalářská práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: RNDr. et PhDr. Ivo Müller Ph.D.

Rok obhajoby práce: 2017

Abstrakt: Hlavním tématem této bakalářské práce jsou pojmy reliabilita a validita, které tvoří důležitá kritéria pro posouzení kvality psychologických testů. V teoretické části je obsažen úvod do problematiky psychometrie a klasické teorie testů, pomocí níž je sestaven teoretický model reliability. Spolu s předpoklady jsou dále představeny metody, kterými míru reliability odhadujeme. Rovněž je vysvětlen pojem validity a přístupy k jejímu určení. V praktické části jsou představeny jednotlivé dotazníky, na kterých byly provedeny výpočty koeficientů reliability dle jednotlivých metod. Závěrečná diskuze je zaměřena na interpretaci výsledků, srovnání hodnot a možnosti užití jednotlivých metod.

Klíčová slova: reliabilita, validita, teorie testů, psychologické testování, dotazníky

Počet stran: 82

Počet příloh: 0

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Michal Bechný

Title: Reliability and validity - different concepts of quality in psychological tests

Type of thesis: Bachelor's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: RNDr. et PhDr. Ivo Müller Ph.D.

The year of presentation: 2017

Abstract: The main topics of the thesis are the concepts of reliability and validity, which make an important criterion for assessing quality of psychological tests. Theoretical part includes an introduction to psychometry and classical test theory by means of which a theoretical model of reliability is constructed. Along with the assumptions, methods used to estimate the degree of reliability are presented. Furthermore, concept of validity and approaches to its determination are explained. Practical part includes and introduces individual questionnaires used for calculating the coefficients of reliability according to individual methods. Final discussion is focused on interpretation of results, comparison of values and possibilities of using the individual methods.

Key words: reliability, validity, test theory, psychological testing, questionnaires

Number of pages: 82

Number of appendices: 0

Language: Czech

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana RNDr. et PhDr. Iva Müllera Ph.D. a všechny použité zdroje jsem uvedl v seznamu literatury.

V Olomouci dne

.....

podpis

Obsah

Úvod	8
1 Psychometrie, její přístupy a teorie testů	9
1.1 Psychologické testování a testy	10
1.2 Klasická teorie testů	10
1.2.1 Model CCT	12
1.3 Teorie položkových odpovědí	12
1.4 Typy položek a jejich kódování	13
1.4.1 Hrubé skóre a transformované skóre	15
2 Reliabilita	16
2.1 Zavedení pojmu	16
2.2 Teoretický model reliability	17
2.3 Metody výpočtu reliability	19
2.3.1 Metoda test-retest	20
2.3.2 Metoda paralelních forem	21
2.3.3 Metoda split half	22
2.3.4 Princip vnitřní konzistence	23
2.4 Vhodná výše reliability	28
3 Validita	29
3.1 Představení myšlenek validity	29
3.2 Přístupy k určení validity	30
3.2.1 Obsahová validita	30
3.2.2 Kriteriaální validita	31
3.2.3 Konstruktová validita	32
3.3 Vhodná výše validity	32
3.4 Srovnání a účel reliability a validity	33

4	Dotazníky	35
4.1	Dotazníky sestavené pro metodu test-retest	36
4.1.1	Dotazník osobního štěstí	36
4.1.2	Test matematických dovedností	38
4.2	Dotazníky sestavené pro metodu paralelních forem	40
4.2.1	Dotazník projevů nervozity při veřejném vystupování	40
4.2.2	Test matematických dovedností	42
4.3	Dotazníky sestavené pro metodu split half	43
4.3.1	Dotazník projevu nervozity při veřejném vystupování	43
4.3.2	Test matematických dovedností	44
5	Představení výpočtů	45
5.1	Dotazník osobního štěstí ($stesti^{TR}$)	46
5.2	Test matematických dovedností (M^{TR})	48
5.3	Dotazník projevů nervozity (N^{PF})	50
5.4	Test matematických dovedností (M^{PF})	52
5.5	Dotazník projevů nervozity (N^{SH})	54
5.6	Test matematických dovedností (M^{SH})	56
6	Diskuze	58
6.1	Srovnání koeficientů reliability	58
6.2	Srovnání hodnot kritériální validity	62
6.3	Shrnutí	63
	Závěr	66

Poděkování

Chtěl bych poděkovat všem svým známým, kteří mě během studia ať už přímo či nepřímo podporovali. Zvláště pak své rodině a panu RNDr. et PhDr. Ivovi Müllerovi Ph.D. za věcné konzultace a inspirativní připomínky.

Úvod

„Pokud člověk dokáže změřit to, o čem mluví, a vyjádřit to číslem, pak o tom něco ví. Je-li to však nemožné vyjádřit číslem, jeho vědomost je skrovná a neuspokojivá. Může to být počátek vědomosti, ale v myšlenkách sotva dosáhl stavu vědy.“

William Thomson - lord Kelvin

Problematika měření je s námi asi tak dlouho jako jsou nám známa čísla. Ta používáme ke kvantifikaci čehokoliv, abychom od stavu neurčitosti došli ke konkrétním představám o míře či intenzitě dané vlastnosti. S čísly jsme již spjati natolik blízko, že jsou pro nás v případě kvantifikace často nositeli srozumitelnější informace nežli slova.

Jedním ze způsobů, kterým se můžeme pokusit změřit vlastnosti či dovednosti člověka je psychologický test. V této práci si představíme, co vše může obnášet jeho konstrukci, zpracování a vyhodnocení. Seznámíme se také s tím, jakými způsoby lze kvantifikovat kvalitu psychologického testu, a sice pomocí dvou důležitých kritérií, kterými jsou reliabilita a validita. S těmito pojmy, které budou tvořit hlavní předmět této práce, se seznámíme jak z teoretického, tak i z praktického hlediska.

Věříme, že tato práce podá čitateli ucelený výklad o problematice psychologického testování a bude užitečným nástrojem při studiu psychometrie a blízkých odvětví.

Kapitola 1

Psychometrie, její přístupy a teorie testů

Známe-li odvěsny pravoúhlého trojúhelníku a chceme spočítat přeponu, ihned si vzpomeneme na známou Pythagorovu větu a po dosazení zjistíme výsledek. Známe přesný vztah, jak se mezi sebou tyto proměnné chovají, a nedělá nám sebemenší problém získat přesný výsledek. Jiná situace nastane, když se nepohybujeme na poli vědního oboru, kde jsou vztahy deterministické a přesně měřitelné. Chceme-li měřit vlastnosti člověka, míru jeho štěstí či spokojenost; pak se budeme zabývat metodami a nástroji, který úzce souvisí s *psychometrií*.

Psychometrie (nebo též psychometrika) je pomocná věda psychologická a zabývá se problematikou měření v psychologii. Hlavními cíli měření může být určování osobnostních charakteristik, zjišťování dovedností, postoju a názorů, diagnostika duševních poruch a další. Jedná se o interdisciplinární obor, který má blízko k psychodiagnostice, osobnostní psychologii a část jeho teoretického pozadí je podepřena matematickou statistikou. Ta dává humanitněji orientované problematice techničtější ráz doplněný o funkční aparát, který slouží ke kvantifikaci a dává nám konkrétní představu o povaze problému.

Zrod psychometrie nastal počátkem 19. století, kdy se problematika učení o lidstvu a společnosti začala zaměřovat i na jedince samotného. Počátky psychometrie lze nalézt v dílech autorů jako např. Francis Galton, Wilhelm Wundt, Alfred Binet (proslavený zkoumáním IQ) či Karl Pearson. Tato zvučná jména známá

i z jiných vědních oborů nám potvrzují myšlenku, že psychometrie je oborem prolínajícím více odvětví.

V rámci psychometrie se vyvinula disciplína zabývající se konstrukcí psychologických dotazníků a testů, která nese patřiný název *teorie testů*. Pod analogicky znějícím anglickým pojmem *Test Theory* se rozumí odvětví, které zkoumá jak teoretické předpoklady nutné pro konstrukci testů, tak i samotný proces tvorby a vyhodnocování psychologických testů. V rámci teorie testů se vžily dva základní přístupy, které si představíme v následujících podkapitolách. Jsou jimi *klasická teorie testů* a *teorie položkových odpovědí* [3]. V obou teoriích je široce uplatnitelná teorie pravděpodobnosti a matematická statistika. Než si tyto dvě teorie představíme, vysvětlíme si, co přesně budeme chápat pod pojmy psychologického testování a testu.

1.1. Psychologické testování a testy

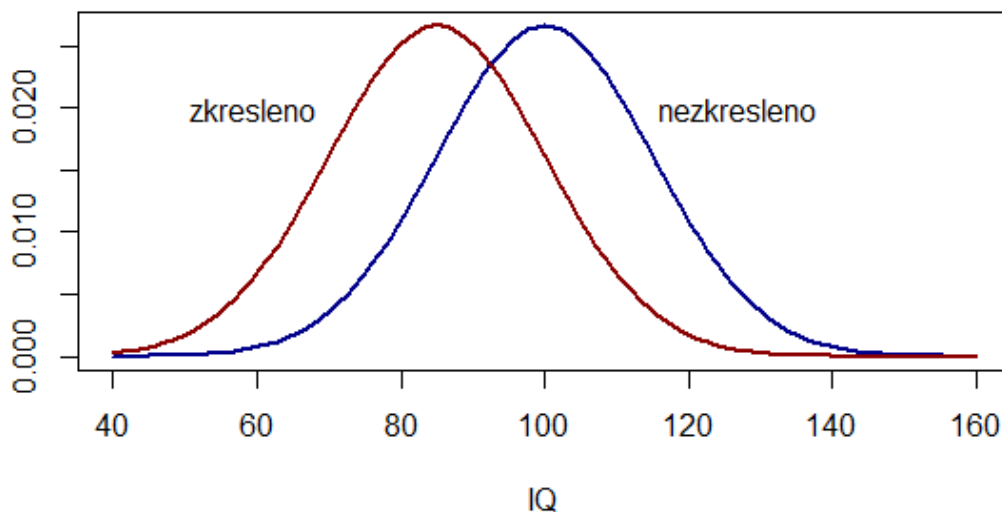
V úvodu jsme si nastínili, jaké mohou být hlavní cíle a motivy psychometrie. Jsou jimi například určování osobnostních charakteristik, zjišťování vlastností či schopností respondenta. Míru, s jakou se tyto atributy u respondenta vyskytují, zjišťujeme pomocí psychologických testů/dotazníků. Ty se skládají z pečlivě vybraných otázek či položek, jejichž vyhodnocením dostaneme informaci o míře nebo intenzitě zastoupení zkoumané vlastnosti respondenta. Proces od vzniku myšlenky přes sestavení testu, účast respondentů, administraci až po samotné vyhodnocení testu nazýváme psychologickým testováním. To může být na základě různých kritérií členěno do mnoha kategorií. Nejzákladnější dělení je na testování: individuální a skupinové; psané a verbální; situační a výkonové; vědomostní, inteligenční a behaviorální[2].

1.2. Klasická teorie testů

Tato teorie, dále uváděná pod zkratkou CTT (z angl. *Classical Test Theory*), bývá ztotožňována s *teorií pravého skóre* (*True Score Theory*)[5, 2]. Označování

teorie jako klasické je z historických důvodů, jelikož vznikla jako první testová teorie. Hlavní myšlenkou je existence pravého (nebo též skutečného) skóre, které je částečně „zastíněno“ chybovou složkou. Pravým skórem je myšlena hodnota, kterou bychom naměřili při naprosté bezchybnosti měření, což je v praxi nereálné. Bezchybností bychom rozuměli naprostou bezchybnost jak testu, tak testovacích podmínek. Chybová složka měření se skládá z náhodné a systematické chyby. Pod systematickou chybou si můžeme představit vliv, který opakovaně ovlivňuje, ať už pozitivně či negativně, skóre, které po vyhodnocení obdržíme. Příkladem může být pískání vzduchotechniky, které zpravidla respondentům zhoršuje koncentraci a tím i snižuje výsledek testu.

Na obrázku 2.1 níže je srovnání nezkráceného a zkráceného testování se systematickou chybou na obrázku (1.1) níže. V normální dětské populaci se předpokládá IQ s průměrnou výší 100 a směrodatnou odchylkou 15. Pokud by testování proběhlo za rušivých vlivů, mohly by být výsledky negativně zkrácené, jak je na obrázku 1.1. Rozptyl by přitom v tomto ukázkovém případě zůstal zachován.



Obrázek 1.1: Jak systematická chyba může ovlivňovat dosažené výsledky

Pod náhodnou chybou si můžeme představit nesystematické působení vnějších

i vnitřních elementů a jiných, systematicky nezachytitelných aspektů. Příklady mohou být chyby způsobené administrací, neobvyklé duševní rozpoložení respondenta (např. nemoc, stres, euforie) nebo vydýchaný vzduch v místnosti. V našem modelu reliability pak nebudeme uvažovat systematickou, ale pouze náhodnou chybu měření, kterou v dalším textu budeme označovat jen jako *chybu měření*. Stejně tak předpokládáme, že pravé skóre a chyba měření jsou navzájem nezávislé náhodné veličiny. U chyby měření budeme dále předpokládat, že má nulovou střední hodnotu a konstantní rozptyl.

1.2.1. Model CCT

Při testování tedy obdržíme *pozorované skóre (observed score)*, které se v rámci našeho modelu skládá ze součtu *pravého skóre (true score)* a chyby měření (*error of measurement*). Jelikož je pozorované skóre tvořeno součtem dvou náhodných veličin, bude se jednat rovněž o náhodnou veličinu. Při vyhodnocování testu *i*-tého respondenta bychom výše uvedený vztah zapsali jako

$$X_i = T_i + \varepsilon_i, \quad (1.1)$$

kde X_i reprezentuje pozorované (nebo též hrubé) skóre, T_i pravé skóre a ε_i chybu měření. Pozorované skóre je možné rozepsat jako

$$X_i = \sum_{j=1}^p X_{ij} \quad (1.2)$$

a interpretovat jej jako součet skóre p -položek v testu vyplněném *i*-tým respondentem. [5]

Podílu variability pravého a pozorovaného skóre se budeme věnovat v dalších kapitolách, kde si představíme koncept *reliability*.

1.3. Teorie položkových odpovědí

Předmětem této teorie, v anglické psané literatuře uváděné jako IRT (*Item Response Theory*), je obdobná problematika jako u CTT, a to sestavování co nejlepšího testu. Teorie položkových odpovědí přichází s řešením pomocí modelování

pravděpodobnosti se kterou respondent označí správnou odpověď u položky, typicky v rámci výkonových testů. Tato pravděpodobnost je modelována jako matematická funkce, která je závislá na přítomnosti *latentních znaků* a *parametrech* položky[1]. Latentním znakem rozumíme charakteristiku, jejíž přítomnost a míru se pomocí testu snažíme vyhodnotit — např. inteligenci, verbální schopnosti, přítomnost poruchy apod.

Pod parametry si můžeme představit například obtížnost otázky. Jiným parametrem může být *diskriminabilita*, tj. míra, s jakou obsah otázky koreluje se zkoumaným latentním znakem. Model této teorie má oproti CTT přísnější předpoklady a jeho matematické pozadí je složitější. Pro účely této práce, kde se budeme zabývat primárně problematikou reliability a validity, budeme vycházet z modelu CTT.

1.4. Typy položek a jejich kódování

Již v průběhu sestavování psychologického nebo i jiného testu je dobré si rozmyslet, jakým způsobem budeme ve fázi vyhodnocování výsledků zpracovávat získané odpovědi. Pro lepší manipulaci se jednotlivé odpovědi kódují pomocí vhodných pravidel na čísla, která představují skóre jednotlivých položek. Součtem skóre jednotlivých položek obdržíme celkové (hrubé) skóre získané respondentem v testu. To lze následně vhodnými úpravami dále transformovat (viz níže).

Pro názornost si v této podkapitole uvedeme několik základních typů položek, které se v psychometrické praxi užívají. K jednotlivým typům je uveden i ilustrační příklad, jakým způsobem je možné odpovědi překódovat na čísla. Většina těchto postupů byla využita pro praktickou část této práce. Pro detailnější informace o plánování experimentu v sociálních vědách se odkazujeme na užitou literaturu[11, 14].

a) Dichotomické položky:

- tazatel má na výběr z možností *ano/ne*,
- typicky kódujeme jako 1/0;

b) trichotomické položky:

- tazatel má na výběr z možností *ano/nevím/ne*,
- kódujeme např. 1/0/-1;

c) výběr jedné správné možnosti:

- typické u výkonnostních testů, např. test znalostí, verbálních dovedností,
- správnou odpověď kódujeme 1 a nesprávnou 0;

d) výběr vícero správných možností (*multiple-choice question*):

- typické pro náročnější a komplexnější výkonnostní testy,
- za každou správnou odpověď můžeme připsat 1 bod,
- výstupem položky je celočíselné ohodnocení;

e) označení možností na škále:

- tato možnost se používá v případě, že chceme znát míru souhlasu s tvrzením či situací,
- škála může být slovní i číselná,
- pro položky musíme mít jednotný škálovací systém, který je třeba v případě reverzní položky invertovat,
- příkladem může být *Likertova škála* složená typicky z možností: *souhlasím, spíše souhlasím, nevím, spíše nesouhlasím, nesouhlasím*,
- kódujeme např. 1,2,3,4,5;

f) sémantický diferenciál:

- typicky pěti- nebo sedmibodová škála je započata a ukončena bipolárními hodnotami, mezi kterými si tazatel volí vhodnou variantu;
- například respondent má na stupnici označit své stravovací návyky, přičemž 1 značí, že se stravuje výhradně zeleninou, a 7 výhradně masem;
- čísla 1,2,3,4,5,6 reprezentují, nakolik je respondent nakloněn té či oné variantě;
- odpovědi kódujeme např. celočíselně.

1.4.1. Hrubé skóre a transformované skóre

Z kódovaných odpovědí můžeme odvodit ukazatele, které nám dávají dobrou představu pro celkové vyhodnocení. Výsledky můžeme prezentovat jak formou hrubého skóre, tak pomocí odvozených charakteristik, které vzniknou transformací skóre hrubého. Hrubým skóre v kontextu testování rozumíme hodnotu přímo naměřeného pozorování. Nejčastěji se tedy jedná o počet správně zodpovězených otázek nebo o součet bodů (kódovaných odpovědí). Hrubé skóre nese kvalitní informaci o absolutní hodnotě výsledku, nikoli však o umístění/srovnání výsledku v rámci testované populace.

Pro tento účel je v praxi zavedeno více odvozených charakteristik, ze kterých si ilustrativně uvedeme dvě. První je percentil, známý zejména z vyhodnocování přijímacího řízení. Dosáhne-li participant výsledku 87. percentil, znamená to, že jeho skóre je lepší nebo stejné pro 87% testovaných a zároveň horší pro zbývajících 13% testovaných.

Jako druhý příklad si uvedeme *T-skóre*, které je velice populární zejména v psychologických kruzích. Jedná se o standardizované skóre s průměrem 50 a směrodatnou odchylkou 10. Pokud budeme předpokládat, že výsledky testování spadají do rodiny normálního rozdělení, pak zhruba 2/3 testovaných obdrží výsledek v rozmezí 40-60. Vyplývá to z pravidla „68-95-99,7“, užívaného v matematické statistice. Toto pravidlo nám referuje o tom, kolik procent pozorování je za výše uvedeného předpokladu ve vzdálenostech menší než je jedna, resp. dvě či tři, směrodatné odchylky od průměru.

Odvozené charakteristiky, jako jsou percentil či T-skóre, neslouží k posouzení kvality testu, ale ke srovnávání v rámci populace respondentů. Pro účely této práce a zjišťování reliability nám tedy bude postačovat znalost hrubého skóre.

Kapitola 2

Reliabilita

2.1. Zavedení pojmu

V předchozích kapitolách jsme se seznámili s různými přístupy psychometrie a s nástroji na vyhodnocování psychologického testu či dotazníku. Kromě obdrženého výsledku nás může zajímat i to, jak kvalitně je test či dotazník sestaven a s jakou spolehlivostí (důvěryhodností) můžeme pohlížet na obdržené skóre. Spolehlivost (angl. *reliability*) v kontextu psychometrie budeme nazývat *reliabilitou*. Jde o kritérium, které bude tím vyšší, čím méně chyb se dopustíme v měření a průběhu testování.

Na reliabilitu lze také pohlížet jako na míru stability výsledků v čase. Tou se rozumí opakovatelnost výsledků, což je míra, s jakou by si byly podobné (blízké) skóre při opětovném testování. Opakovatelnost i celková spolehlivost testu bude tím vyšší, čím méně bude testování zatíženo chybou měření, náhodnými fluktuacemi a celkově špatnou konstrukcí a formulací otázek. V dalším textu se seznámíme jak s modelem, který nám odhalí teoretické pozadí reliability, tak i s přesnou definicí. V praktické části si představíme jednotlivé postupy výpočtů a budeme je aplikovat na reálných dotaznících a testech, které byly sestaveny pro potřeby této práce.

2.2. Teoretický model reliability

Pokud bychom zaslechli tvrzení, že daný psychologický test je důvěryhodný, zřejmě bychom se spokojili pouze částečně. Rozsah této důvěryhodnosti je proto nezbytné kvantifikovat, abychom si o její míře mohli udělat konkrétní představu. Pro tyto účely si na základě klasické teorie testů odvodíme a zadefinujeme pojem *reliability*, převzatý z literatury[1, 2].

Vyjdeme ze dříve zavedeného vztahu (1.1) pro pozorované skóre i -tého subjektu: $X_i = T_i + \varepsilon_i$, který uvažujeme pro všechna $i = 1, \dots, n$. Pro fungování modelu reliability budeme uvažovat nezávislost všech měření X_i a další důležité předpoklady:

- střední hodnota chyby měření je nulová,

$$E(\varepsilon_i) = 0; \quad (2.1)$$

- rozptyl chybových složek měření je konstantní,

$$\text{var}(\varepsilon_i) = \sigma^2; \quad (2.2)$$

- chyby měření jsou mezi sebou nezávislé,

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0; \quad (2.3)$$

- veličiny pravé skóre a chyba měření jsou nekorelovány,

$$\text{cov}(T_i, \varepsilon_i) = 0. \quad (2.4)$$

Pro rozptyl pozorovaného skóre pak podle vztahu pro nekorelované náhodné veličiny musí platit rovnost

$$\sigma_X^2 = \sigma_T^2 + \sigma_\varepsilon^2. \quad (2.5)$$

Z výše uvedených předpokladů a vztahu pro rozptyl pozorovaného skóre snadno přejdeme k definici reliability.

Definice 1 *Reliabilitou* budeme rozumět reálné číslo ρ_{xx} vyjádřené jako podíl rozptylu pravého skóre na celkovém rozptylu.

Lze tedy zapsat

$$\rho_{xx} = \frac{\sigma_T^2}{\sigma_X^2}. \quad (2.6)$$

Stejně tak můžeme reliabilitu chápat jako reálné číslo ρ_{xx} , vzniklé tak, že od jedničky odečteme podíl rozptylu chybové složky na celkovém rozptylu v obdržených datech. Jak ukáže matematický zápis, jedná se o totožný výraz

$$\rho_{xx} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_X^2 - \sigma_\varepsilon^2}{\sigma_X^2} = 1 - \frac{\sigma_\varepsilon^2}{\sigma_X^2}. \quad (2.7)$$

Z výše uvedených vztahů je zajištěno, že reliabilita bude vždy nabývat hodnot na intervalu $\langle 0; 1 \rangle$. Tuto vlastnost lze demonstrovat na vztahu (2.7), kde od čísla 1 odečítáme podíl $\frac{\sigma_\varepsilon^2}{\sigma_X^2}$. Z kladnosti rozptylu a vlastnosti $\sigma_\varepsilon^2 < \sigma_X^2$ plyne, že tento rozdíl musí být při nenulovém σ_ε^2 větší než nula. Reliabilita o hodnotě 1 by indikovala perfektní a v praxi nedosažitelnou reliabilitu[1].

V následujících odstavcích si ukážeme, v jakém vztahu jsou pojmy reliabilita a korelace. Předpokládáme, že pojem korelace je čtenáři znám včetně svých vlastností a zopakujeme si pouze její obecné vyjádření pro libovolné náhodné veličiny X a Y. Hodnota korelace $\rho_{X,Y}$, která je teoretickou mírou pro míru lineární závislosti dvou veličin, je dána vztahem

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}}. \quad (2.8)$$

Pro výpočet výběrového korelačního koeficientu $r_{X,Y}$ nahradíme příslušné teoretické charakteristiky výběrovými. Touto úpravou a následným zjednodušením výrazu získáváme tzv. Pearsonův korelační koeficient, který nabývá hodnoty z intervalu $\langle -1, 1 \rangle$ a je ve tvaru

$$r_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (2.9)$$

Nyní si užitím výše popsaného vztahu ukážeme, že lze reliabilitu vyjádřit i jako druhou mocninu korelačního koeficientu mezi pravým a pozorovaným skóre.

$$\rho_{T,X}^2 = \frac{\text{cov}^2(T, X)}{\sigma_T^2 \sigma_X^2} = \frac{\text{cov}^2(T, T + \varepsilon)}{\sigma_T^2 \sigma_X^2} = \frac{(\sigma_T^2)^2}{\sigma_T^2 \sigma_X^2} = \frac{\sigma_T^2}{\sigma_X^2} = \rho_{xx}. \quad (2.10)$$

V prvním kroku jsme zapsali dle vztahu (2.8) druhou mocninu korelace. Dále jsme veličinu pozorovaného skóre rozepsali jako $X = T + \varepsilon$. V důsledku nezávislosti veličin T a ε jsme do další rovnosti v čitateli obdrželi druhou mocninu rozptylu pravého skóre, která se v poslední úpravě vykrátila s rozptylem pravého skóre ve jmenovateli.

Úpravami druhé mocniny korelace jsme obdrželi podíl variability pravého skóre na variabilitě skóre pozorovaného. Je zřejmé, že interpretace tohoto podílu je totožná s definicí reliability. Poznamenejme, že výše popsáný podíl je totožný s indexem determinace, který je definován jako podíl variability vysvětlované modelem (zde pravého skóre) a celkové variability obsažené v pozorovaných datech (zde v pozorovaném skóre). Nyní vidíme jasnou spojitost mezi reliabilitou a korelací.

Reliabilita se dá také vyjádřit jako korelace mezi testy, které měří stejný atribut. Pro pozorované skóre těchto testů platí, že $X = T + \varepsilon$ a $Y = T + \gamma$. Předpokládáme shodnost rozptylů těchto pozorovaných skóre $\sigma_X^2 = \sigma_Y^2$ a také vlastnosti pro chybové složky ε, γ popsané výše. Zřejmě platí, že

$$\text{cor}(X, Y) = \frac{\text{cov}(T + \varepsilon, T + \gamma)}{\sqrt{\sigma_X^2 \sigma_Y^2}} = \frac{\sigma_T^2}{\sigma_X^2} = \rho_{xx}. \quad (2.11)$$

2.3. Metody výpočtu reliability

V praxi však nemáme ani nástroj, ani techniku, jak vypočítat hodnotu variability pravého skóre přímo. Proto si v následující kapitole představíme několik hlavních metod, jak podíl variability pravého skóre na variabilitě pozorovaného skóre odhadnout a tím zároveň odhadnout hodnotu reliability. Těmito metodami budou *test-retest*, *metoda paralelních forem*, *split half* a metody nahlížející na reliabilitu přístupem *vnitřní konzistence*[2]. Každá metoda má své klady a zápory

a zároveň do jisté míry limitující předpoklady.

Abychom ukázali, že při praktických výpočtech teoretickou hodnotu reliability ρ_{xx} odhadujeme, budeme označovat její výběrový odhad značený písmenem r , který nazveme jako *koeficient reliability*. Pro zdůraznění postupu, kterým budeme reliability odhadovat, obohatíme tento výběrový odhad o příslušné horní a dolní indexy.

2.3.1. Metoda test-retest

Název této metody nese zároveň i její hlavní myšlenku, kterou je tedy zachycení stability (opakovatelnosti měření) v čase. To se provede pomocí aplikování téhož dotazníku (testu) na tytéž respondenty s určitým časovým odstupem. Následně se vyhodnocuje, do jaké míry jsou výsledky těchto dvou testování blízké[2]. Představu o míře blízkosti („těsnosti“) mezi těmito dvěma skupinami výsledků nám dobře reprezentuje Pearsonův korelační koeficient popsáný vztahem (2.12), jehož hodnota bude v tomto případě reprezentovat samotný koeficient reliability, který v tomto případě budeme značit r^{TR} . Pro snadnější výpočet můžeme uvedený vztah upravit na tvar

$$r^{TR} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}}, \quad (2.12)$$

kde n představuje počet respondentů, x_i skóre obdržené i -tým respondentem v rámci prvního testu a y_i skóre téhož respondenta v rámci druhého testu (retestu).

Pokud bychom předpokládali neměnnost zkoumaného atributu i chybové složky v čase, obdrželi bychom identické výsledky. Tento předpoklad je však v praxi porušen hned několika způsoby. Dochází jak ke změnám působení chybové složky, která se v souladu s předpoklady chová nahodile, tak ke změnám hodnoty pravého skóre, které je v průběhu času ovlivněno nejen procesem učení, ale i probíhajícími psychickými procesy.

Problém učení se častěji vyskytuje u mladistvých, kde i během krátkého intervalu může dojít k významnějším změnám ve vědomostech, což v důsledku může ovlivnit získané skóre u znalostních testů. Výsledek testování může být také

ovlivněn změnami v chybové složce, pod kterou si můžeme představit rozkolísanost nálady, pocíťovaný stres, ovzduší v místnosti a další nesystematické vlivy, které se v čase chovají rovněž proměnlivě.

Jak vidíme, časový odstup může ovlivňovat jak osobnostní, tak i výkonové testy. Čím menší je interval mezi testováními, tím se mohou obdržené výsledky jevit ekvivalentnějšími („s vyšší reliabilitou“) a obráceně. Při neúměrně krátkém intervalu mezi testováními by mohlo dojít k zapamatování si odpovědí na položky a tím k nežádoucímu nadhodnocení koeficientu reliability. Časový interval se doporučuje volit jako kompromis mezi tím, aby nedošlo k výrazné změně pravého skóre a zároveň se zamezilo zapamatování si odpovědí na jednotlivé položky. Jak vidíme, mezi negativa této metody patří zejména náklady spojené s dvojí administrací a problematika stanovení intervalu mezi testováními.

2.3.2. Metoda paralelních forem

Na tuto metodu se dá nahlížet jako na vhodnou alternativu k metodě test-retest v případě, že se chceme vyhnout problému se stanovením délky intervalu mezi testováními a dvojí administrací. Principem metody paralelních (též ekvivalentních, alternativních) forem je otestování respondentů dvěma dotazníky (testy) v jednom časovém okamžiku. Výstupem administrace jsou dvojice (X_i, Y_i) , představující pozorované skóre i -tého respondenta v rámci jedné a druhé testové formy. Aby námi počítaný odhad koeficientu reliability měl co nejlepší výpovědní hodnotu o její skutečné hodnotě, klademe v rámci této metody na spárované dvojice položek v jednotlivých sadách testů následující požadavky[2, 5]:

- počet položek u obou testů je stejný,
- každá spárovaná dvojice položek je ekvivalentní z hlediska obsahu, vlastností a vyžadovaných operací,
- obtížnost položek je v testových formách identicky rozložena,
- střední hodnoty a rozptyly skóre paralelních testových forem jsou shodné,

- proces administrace a skórování je totožný.

Jelikož je úplné splnění těchto požadavků nedosažitelné, je v praxi požadováno co největší přiblížení se. Koeficient reliability r^{PAR} se stejně jako u předchozí metody stanoví pomocí Pearsonova korelačního koeficientu (2.12). Výhodami metody paralelních forem oproti předchozí metodě jsou administrace provedená v jednom časovém okamžiku, stejné externí i interní podmínky při testování a znemožnění vlivu učení se a pamatování si na výsledky jednotlivých položek.

Díky těmto vlastnostem je metoda paralelních forem velmi hojně využívána. Pro co nejlepší přiblížení se požadavkům je vyžadováno značné úsilí věnované konstrukci dotazníků, které je však vykompenzováno tím, že výpočet reliability touto metodou dává nejkvalitnější výsledky z prezentovaných metod[5].

2.3.3. Metoda split half

Pokud nám okolnosti testování nedovolují zjišťovat reliabilitu formou dvou dotazníků (testů), například kvůli nákladům a pracnosti spojeným s jejich sestavováním, lze využít metodu *split half*. K jejímu užití postačuje jeden dotazník, jehož položky rozdělíme do dvou podskupin (odtud český název *metoda dvou polovin*). Toto rozčlenění se dá provést vícero způsoby.

Jedna z možností je rozdělit položky podle toho, zda spadají do první nebo druhé poloviny testu. Pokud mají položky v testu stoupající náročnost, je vhodnější rozdělení provést tak, že položky budeme seskupovat dle sudosti (lichosti) jejich pořadí. Další možností je čísla položek, a tím i jejich rozřazení, náhodně losovat.

Namísto celkového skóre obdržíme u každého respondenta dvě hodnoty: skóre přiřazené první a druhé skupině položek. Dvojici těchto hodnot pro i -tého respondenta zapíšeme jako $(X_i^{(1)}, X_i^{(2)})$. Koeficient reliability, v tomto případě značený r^{SH} , se vypočte pomocí *Spearmanova-Brownova vzorce*[14, 5]

$$r^{SH} = \frac{2r}{r + 1}, \quad (2.13)$$

kde r značí Pearsonův korelační koeficient mezi veličinami $X^{(1)}$ a $X^{(2)}$, vypočítaný vztahem (2.12).

Spermanův-Brownův vzorec (též korekce), byl v r. 1910 nezávisle odvozen pány Spearmanem a Brownem. Užívá se v obecném tvaru ke zjišťování celkové reliability testu, který byl rozdělen na k stejných částí. Tvar tohoto vzorce, který je i v odborné literatuře nazýván jako věštecký, je obecně

$$r^* = \frac{kr}{1 + (k - 1)r}. \quad (2.14)$$

Položíme-li $k = 2$, dostáváme výše uvedený vztah (2.13) pro výpočet split-half reliability.

2.3.4. Princip vnitřní konzistence

Dalším možným přístupem, jak odhadovat reliability, je nahlížet na každou položku testu jako na samostatný subtest. Předpokládáme, že všechny položky zkoumají tentýž aspekt, jen pomocí jiné formulace. Měří-li položky totéž, dá se předpokládat jejich vzájemná kladná korelace. Ta je pak v tomto kontextu nazývána jako konzistence testu.

Díky těmto vlastnostem se dá reliability odhadnout pomocí principu obdobného jako u metody split-half, kde se test rozdělil na dvě skupiny a reliability se vyjádřila pomocí korelace mezi skóre v obou skupinách. Nyní tuto úvahu zobecníme a rozdělíme test na p podskupin, přičemž je p rovno počtu položek v původním testu. Reliabilitou budeme rozumět míru průměrné korelace všech položek testu.

Výpočetní odvození odhadu reliability tohoto typu, jakožto zobecnění předchozí metody, bylo provedeno v roce 1951 americkým psychologem Leem Josephem Cronbachem v díle *Alpha paper*[13], odkud se později vžil dodnes užívaný název jako Cronbachova alfa. Vztah vznikl syntézou a zobecněním několika metod, které se v té době již používaly (např. KR-20 uvedená níže). Jedná o jednu z nejčastěji užívaných metod ke zjišťování vnitřní konzistence testů a tím i k výpočtu reliability[5, 14, 2].

a) Cronbachův hrubý koeficient alfa

Pro tento obecný tvar Cronbachova alfa budeme uvažovat, že se položky mohou lišit ve své struktuře i možnostech ohodnocení. Nevznášíme požadavek na výskyt správné odpovědi a bude nám postačovat, že každá možnost má své vlastní bodové ohodnocení.

V důsledku uvažujeme různý rozptyl σ_j^2 pro každou z p položek a rozptyl σ_X^2 celkového testového skóre. Zopakujeme, že pozorované skóre u i -tého respondenta je dáno jako $X_i = \sum_{j=1}^p X_{ij}$. Výběrovou hodnotu koeficientu reliability v tomto případě značíme r_α^R a dle [2] spočítáme jako

$$r_\alpha^R = \frac{p}{p-1} \left(1 - \frac{\sum_{j=1}^p S_j^2}{S_X^2} \right). \quad (2.15)$$

Ve vztahu (2.15) jsou již uvažované (teoretické) rozptyly nahrazeny jejich výběrovými hodnotami. Výběrovou hodnotu σ_j^2 reprezentuje výběrový rozptyl j -té položky, který vypočteme jako

$$S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_{.j})^2 \quad \forall j = 1, \dots, p, \quad (2.16)$$

kde X_{ij} představuje skóre i -tého respondenta na j -tou položku, $\bar{X}_{.j}$ značí průměrné skóre získané v j -té položce za všech n respondentů. Analogicky, výběrová hodnota σ_X^2 , rozptylu pozorovaných skóre, bude zastoupena příslušným výběrovým rozptylem ve tvaru

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad (2.17)$$

kde X_i představuje skóre i -tého respondenta v celém testu a \bar{X} představuje příslušné průměrné skóre získané v testu mezi všemi respondenty. Hrubý koeficient alfa lze také vyjádřit užitím vztahu

$$r_\alpha^R = \frac{p\bar{c}}{\bar{v} + (p-1)\bar{c}}, \quad (2.18)$$

kde \bar{c} je průměrná kovariance mezi p položkami a \bar{v} je průměrný rozptyl položek.

Přechod k tomuto zápisu si nyní odvodíme. Užijeme vlastnosti celkového skóre X , které jsme pro každého z n respondentů vyjádřili jako součet skóre obdržených v jednotlivých p položkách. Jednotlivá i celková skóre jsou náhodné veličiny a charakteristiky, které pomocí nich získáme budeme napočítávat vždy přes všechny respondenty, tzn. $i = 1, \dots, n$. Při odvození užijeme vlastnosti rozptylu celkového skóre, které je vlastně rozptyl součtů, což můžeme rozepsat do následující podoby

$$\mathit{var}(X) = \mathit{var}\left(\sum_i \sum_{j=1}^p X_{ij}\right) = \sum_{j=1}^p \sum_{k=1}^p \mathit{cov}_i(X_{ij}, X_{ik}). \quad (2.19)$$

Nyní tuto rovnost nahradíme výběrovými charakteristikami. Rozptyl celkového skóre (2.17) lze tedy rozepsat na součty výběrových kovariancí a rozptylů.

$$S_X^2 = \sum_{j=1}^p \sum_{k=1}^p \widehat{\mathit{cov}}_i(X_{ij}, X_{ik}) = \sum_{j=1}^p \widehat{\mathit{var}}_i X_{ij} + \sum_{j=1}^p \sum_{\substack{k=1 \\ j \neq k}}^p \widehat{\mathit{cov}}_i(X_{ij}, X_{ik}). \quad (2.20)$$

Označíme-li odhad rozptylu j -té položky jako S_j^2 a kovarianci mezi j -tou a k -tou položkou jako c_{jk} , lze tento vztah přepsat na

$$S_X^2 = \sum_{j=1}^p S_j^2 + \sum_{j=1}^p \sum_{\substack{k=1 \\ j \neq k}}^p c_{jk}. \quad (2.21)$$

Označíme-li průměrné hodnoty kovariancí a rozptylů jako

$$\bar{c} = \frac{1}{p(p-1)} \sum_{j=1}^p \sum_{\substack{k=1 \\ j \neq k}}^p c_{jk} \quad (2.22)$$

$$\bar{v} = \frac{1}{p} \sum_{j=1}^p S_j^2 \quad (2.23)$$

můžeme rozptyl celkového skóre ze vztahu (2.21) přepsat na

$$S_X^2 = p\bar{v} + p(p-1)\bar{c}. \quad (2.24)$$

Dosazením vztahů (2.23) a (2.24) do (2.15) získáváme vzorec ve tvaru (2.18)

$$r_{\alpha}^R = \frac{p}{p-1} \left(1 - \frac{p\bar{v}}{p\bar{v} + p(p-1)\bar{c}} \right) = \frac{p}{p-1} \left(\frac{p(p-1)\bar{c}}{p\bar{v} + p(p-1)\bar{c}} \right) = \frac{p\bar{c}}{\bar{v} + (p-1)\bar{c}}. \quad (2.25)$$

b) Cronbachův standardizovaný koeficient alfa

Kdybychom při dosazování hodnot do vztahu (2.15) za pozorovaná skóre respondentů v jednotlivých položkách dosadili jejich standardizované z-skóre, obdrželi bychom tzv. standardizovaný koeficient alfa r_{α}^{st} . Z-transformace se v tomto případě provádí tak, že od každého pozorování odečteme příslušnou střední hodnotu skóre otázky a následně tento rozdíl vydělíme odmocninou z rozptylu dané otázky (tj. směrodatnou odhylkou). V případě výběrových hodnot se uvedený postup z-transformace zapíše jako

$$X'_{ij} = \frac{X_{ij} - \bar{X}_{.j}}{\sqrt{S_{.j}^2}}. \quad (2.26)$$

Další možností vyjádření standardizovaného koeficientu je dosazení nestandardizovaných (hrubých) hodnot do vztahu

$$r_{\alpha}^{st} = \frac{p\bar{r}}{1 + (p-1)\bar{r}}, \quad (2.27)$$

kde \bar{r} značí průměrnou korelaci vypočítanou mezi všemi p položkami. Vztah (2.27) je analogií vztahu (2.18). V tomto vztahu je namísto kovariance užito korelace, která svou konstrukcí data standardizuje. V případě standardizovaných položek je kovariance rovná korelaci a rozptyl je roven jedné, což platí i pro průměrné hodnoty.

Chceme-li tedy vypočítat standardizovaný koeficient alfa, máme dvě možnosti. Buď můžeme dosadit standardizované položky do vztahu (2.18) nebo použít vztah (2.27), který svou povahou položky standardizuje.

Standardizovaný koeficient alfa má smysl uvažovat zejména v případech testů, jejichž položky mají různé hodnotící a bodové škály. Standardizací tyto vlivy

eliminujeme a veškeré položky převádíme na stejnou jednotku (směrodatnou odchylku). Pokud jsou položky testu standardizovány, je použití tohoto koeficientu diskutabilní.

c) Koeficient KR-20

Ve výkonových testech bývá častým jevem, že pro každou položku existuje právě jedna správná odpověď. Tu můžeme kódovat číslem 1 a naproti tomu špatnou odpověď číslem 0. V případě jediné správné odpovědi pro každou z p položek má smysl uvažovat index (koeficient) obtížnosti, tzv. *difficulty index*. Pro j -tou položku jej budeme značit DI_j . Tento index nám stanovuje, jaká proporce respondentů na danou položku odpověděla správně, respektive špatně.

Jelikož nemůže nastat žádný jiný jev, bude mít správnost odpovědi respondenta na j -tou otázku alternativní rozdělení pravděpodobností s parametrem p_j . V kontextu alternativního rozdělení značí parametr p_j pravděpodobnost úspěchu, kterým v tomto případě rozumíme správné zodpovězení j -té otázky. Jeho doplněk, který označíme q_j , pak bude označovat pravděpodobnost označení chybné, tedy jakékoli jiné než správné odpovědi[2]. Je zřejmé, že $q_j = 1 - p_j$. Provedeme-li dále součin $p_j q_j$, získáme rozptyl j -té položky.

Výběrové odhady parametrů p_j, q_j získáme z obdržených dat jako proporce správných, resp. špatných, odpovědí. Je zřejmé, že $\widehat{p}_j = DI_j$. Princip i výpočet reliability jsou totožné s Cronbachovým alfa. V tomto případě je vzorec pro výpočet koeficientu reliability r^{KR} ve tvaru

$$r^{KR} = \frac{p}{p-1} \left(1 - \frac{\sum_{j=1}^p DI_j(1 - DI_j)}{S_X^2} \right), \quad (2.28)$$

kde S_X^2 je výběrový rozptyl obdržených celkových skóre popsany vztahem (2.17).

Vztah (2.28) bývá nazýván jako Kuderův-Richardsonův vzorec číslo 20. Poukazuje na jména autorů, kteří jej v roce 1937 odvodili[14]. Číslo 20 pak připomíná číslování, které v článku náleželo právě tomuto vzorci. Název se ujal a hojně se užívá dodnes.

2.4. Vhodná výše reliability

Jak je nám známo z teoretického modelu, reliability je omezena čísly 0 a 1. Obecně platí, že čím více je námi zjištěný koeficient blíže 1, tím vyšší jsme obdrželi reliability. Je dosti subjektivní, jaká výše reliability je dostatečná, a proto se budeme řídit obecně uznávanými standarty[8] pro interpretaci kvality korelačního koeficientu.

Koeficient reliability	Interpretace
$r \geq 90$	výborná
$0.80 \leq r < 0.90$	dobrá
$0.70 \leq r < 0.80$	uspokojivá
$0.70 < r$	diskutabilní

Požadovaná výše koeficientu reliability se také liší také dle oblasti použití. Příklady jsou uvedeny v následující tabulce.

Testovaný atribut	Požadovaná reliability
IQ	$r \geq 90$
osobnost	$r \geq 70$
kreativita	$r \geq 50$

Kapitola 3

Validita

3.1. Představení myšlenek validity

Dalším z ukazatelů kvality testů je *validita*, která nám odpovídá na otázku, jak moc test měří to, co měřit má. Jinými slovy, je to stupeň, se kterým nám výsledky testu referují o skutečné hodnotě testované veličiny. Neboli jak moc koresponduje výsledek testu s jinými nezávislými posudky považovanými za objektivní a mající dobrou vypovídající hodnotu[6].

V kontextu psychologického testování se pod pojmem validita rozumí přesnost, s jakou můžeme pohlížet na objektivitu obdržených skóre. Proneseme-li v českém jazyce výrok, že je něco *validní*, znamená to platný. Tento ekvivalent nemá daleko ani od skutečného významu validity v kontextu psychologického testování.

Pokud bychom validitu uvažovali pohledem statistiky, hovořili bychom o tzv. nevychýlenosti. Měření či odhad parametru se považuje za nevychýlené (též ne-stranné), pokud se jeho střední hodnota rovná právě hodnotě parametru, který se snaží odhadnout. Napříč těmito příklady z různých oborů vidíme, že měření (psychologické testování, odhad) či výrok je tím validnější, čím blíže je jeho výsledek skutečné (pravé) hodnotě.

Validita je kritérium, jehož hodnocení by nemělo chybět u žádné psychologické studie či testování. Při sestavování testu je určitě vhodné, vedle zohledňování peněžních nákladů a etické stránky testování, věnovat i patřičný důraz validitě, která dá jasnou představu o tom, zda námi používaný dotazník (test) měří

skutečně to, pro co je určen.

Z výše uvedených úvah je zjevné, že se validita vyskytuje napříč odvětvími, a zároveň ji každé z těchto odvětví chápe a vysvětluje s drobnými odlišnostmi. Samotná validita je oproti reliabilitě hůře definovatelná a ještě obtížněji kvantifikovatelná. Na následujících stránkách této práce si validitu představíme v kontextu psychometrie. Při zpracování této kapitoly bylo čerpáno ze zdrojů [2, 5, 6, 8, 10, 12].

3.2. Přístupy k určení validity

Pojem validity je značně abstraktní a není zcela přesně vymezen či definován. Jako důsledek validace pro praktické účely, v průběhu vývoje této relativně rané problematiky, vykrytalizovaly jednotlivé přístupy, jak na validitu nahlížet a případně ji i vypočítat. První myšlenky a praktické výpočty validity probíhaly v USA za druhé světové války, kdy docházelo k obrovskému rozmachu psychometrie, jako důsledku vysoké poptávky ze strany armády. Bylo potřeba zajistit dobrou validitu testů např. pro nábor leteckých pilotů, kteří museli disponovat značnou psychickou odolností a vybranými osobnostními atributy. V následující kapitole si představíme základní přístupy, kterými se dá na validitu nahlížet z různých hledisek.

3.2.1. Obsahová validita

Tato validita, v anglické literatuře známá jako *content validity*, se užívá hlavně u testů výkonnosti, kde zjišťujeme úroveň, se kterou testovaný ovládá danou problematiku či schopnost. Myšlenkou je vybrat co nejrepresentativnější okruh testových položek napříč celou oblastí týkající se testovaného atributu, abychom mohli co nejobjektivněji posoudit rozsah znalostí respondenta. Reprezentativní výběr položek je často dosahován konzultací s experty na danou problematiku.

Jako příklad obsahové validity si můžeme představit test z vysokoškolské matematiky. Pedagog musí pro test vybrat takové úlohy, které budou dobře reprezentovat požadované znalosti a zároveň pokryjí celé spektrum zkoušené proble-

matiky. Pokud by položky takovými nebyly, výsledky testování by v rámci dané situace nebyly validními a nedávaly by nám dobrou představu o schopnostech respondenta.

Face Validity

Méně striktním a spíše pomocným přístupem pro konstrukci testů je koncept *face validity*. Myšlenka je taková, že zkonstruovaný test je předložen nezávislé osobě, která po letmém prohlédnutí položek vyhodnotí, zda je test dobře konstruovaný (=validní) a je sestaven tak, aby se obsahem svých položek na první pohled vztahoval k testovanému atributu, kterým může být jak znalost, tak i osobnostní rys. Tato „validace nahlédnutím“ (odkud název *face validity*) určitě není úplně objektivním postupem, a proto se užívá zejména jako podklad pro pilotní studie.

3.2.2. Kriteriaální validita

Tento přístup stanovuje míru validity pomocí srovnání výsledku testového skóre s jiným, objektivním kritériem, které zkoumá stejný aspekt a je na našem testování nezávislé. Příkladem může být testování inteligence, kde se v dřívějších dobách výsledek validoval na školním prospěchu, který sehrával roli kritéria. Míru validity, v tomto případě toho, že je námi obdržené skóre v blízkém vztahu s objektivním kritériem, budeme stanovovat pomocí Pearsonova korelačního koeficientu. Budeme rozlišovat dva základní druhy kriteriaální validity.

Pokud u respondenta budeme v jeden moment zkoumaný aspekt testovat pomocí dvou testů, z nichž byl jeden již dříve validovaný, lze zkoumat míru jejich vzájemné korelace. Dříve validovaný test slouží jako validační kritérium a celému přístupu s touto myšlenkou říkáme *souběžná validita*. Aplikujeme-li tyto dva testy s určitým časovým odstupem, lze uvažovat, že skóre prvního testu bude mít prediktivní hodnotu o výsledku druhého testování. Tento postup souhrnně označujeme jako *prediktivní validaci*. Jsou použity stejné testy jako u předchozí metody, avšak s časovým odstupem.

V obou případech je validita vyčíslena korelačním koeficientem. Pokud bychom tento koeficient umocnili, obdrželi bychom index determinace, známý z regresní analýzy a obdobný jako ve vztahu (2.10). Ten by nám dával informaci o tom, jaká část variability v kritériu je vysvětlena testem.

3.2.3. Konstruktová validita

Přístup konstruktové validity oproti těm předchozím je daleko komplexnější a zahrnuje v sobě více proměnných. Proces validace se neuskutečňuje jen pomocí kritérií, které se zkoumaným aspektem přímo souvisí, ale také pomocí těch, které s aspektem nesouvisí. Myšlenka je taková, že kritéria blízká tomu zkoumanému by měla být korelována více (a kladně), kdežto kritéria, která jsou v rozporu se zkoumaným atributem, by měla být teoreticky korelována slabě, nulově nebo záporně. Jako příklad si představme, že budeme zkoumat extroverzi lidí. Měla by být kladně korelována s navazováním nových kontaktů a mezilidskou komunikací, a zároveň velmi negativně korelována například s potřebou samoty, výskytu úzkosti apod. O kladné a silné korelaci s příbuznými kritérii se hovoří jako o *konvergentní validitě*, naopak u slabé či záporné jako o *divergentní validitě*.

3.3. Vhodná výše validity

Obdobně jako v případě reliability bude obtížné stanovit optimální, resp. minimální požadovanou, výši validity. V této kapitole jsme si představili základní přístupy k určování validity. Po pročtení je zjevné, že její výpočet je možné uvažovat pouze pro případ kriteriálního typu, kde uvažujeme korelaci mezi kritériem a obdrženým testovým skóre. Hodnotu korelačního koeficientu v tomto případě budeme označovat jako *validační koeficient*. Pro vyhodnocení kvality obdržené validity se pro tento případ budeme řídit obecně uznávanými standarty[8], které jsou uvedeny v následující tabulce:

Hodnota validačního koeficientu	Interpretace
> 0.35	velmi dobré
$0.21 - 0.35$	postačující
$0.11 - 0.20$	diskutabilní
< 0.10	nedostatečné

V ostatních případech se hodnoty validity nedají nikterak číselně vyjádřit a dochází k jejímu subjektivnímu hodnocení. V důsledku této přítomnosti subjektivity dochází v literatuře i napříč odborníky k odlišnému definování validity a chápání její podstaty. Téma validity nejen v kontextu psychologického testování, ale i veškerého vědeckého výzkumu je aktuální a pro zájemce o hlubší problematiku odkazujeme na seznam použité literatury.

3.4. Srovnání a účel reliability a validity

Reliabilita a validita jsou charakteristiky, které mají společné to, že vypovídají o kvalitě testu. Hledisko, podle kterého kvalitu posuzují, je však rozlišné. Často se kombinace jejich možných vztahů ilustruje na příkladu střelby na terč, kde mohou nastat čtyři základní situace mezi výsledky. Reliabilita je ukazatelem stability a opakovatelnosti výsledků, zatímco validita posuzuje, nakolik jsou výsledky platné a správné. Zjištěním a vyhodnocením reliability a validity dostáváme představu



Obrázek 3.1: Možné vztahy reliability a validity

o tom, nakolik je chyba v testování systematická či náhodná, a tím i informaci, zda je měření zatíženo určitým vychýlením. Pro praktické experimenty je důležitá

vysoká míra kvality obou charakteristik - jedině tak bude dosaženo relevantních výsledků, které budou objektivní a experimentálně reprodukovatelné.

Kapitola 4

Dotazníky

V minulých kapitolách jsme se seznámili s oborem psychometrie, jejími přístupy a také jsme si představili úvod do problematiky psychologického testování. Pro ohodnocení kvality, se kterou testování probíhá, jsme si zavedli pojmy reliability a validity. Nyní přejdeme k praktické části této práce, která primárně spočívala v sestavení a vyhodnocení testů uzpůsobených pro výpočet koeficientu reliability vždy dle dané metodiky.

Představíme si celkem šest rozdílných testů, které jsou uskupeny do tří dvojic. Každý pár obsahuje jeden osobnostní a jeden výkonnostně orientovaný test. Tyto tři páry jsou svou formou, v souladu s předpoklady, uzpůsobeny k výpočtu koeficientu reliability třemi metodami, tedy postupně metodou test-retest, paralelních forem a split half.

V rámci osobnostních testů byl použit dotazník osobního štěstí a také dva rozlišné dotazníky zkoumající projevy nervozity při veřejném vystupování. Dotazník osobního štěstí je v rámci šestice použitých testů jediným převzatým dotazníkem. Zbývajících pět testů bylo pro účely této práce vlastnoručně sestrojeno. V rámci výkonnostně orientovaných testů byly pro jednotlivé metody sestaveny tři rozlišné testy zkoumající matematické dovednosti. Jelikož výpočty metodami vnitřní konzistence nekladou na formu testu žádné zvláštní požadavky, byly provedeny na již představených dotaznících.

Nyní se seznámíme s jednotlivými dotazníky/testy, jejich respondenty a procesem, kterým byly administrovány. V další kapitole si pak shrneme výpočty a v

diskuzi se zamyslíme nad získanými výsledky a interpretací. Všechny sestavené dotazníky jsou součástí přílohy. U každého dotazníku uvedeme v závorce zkratku či heslo, pod kterým jej dále uvádíme.

4.1. Dotazníky sestavené pro metodu test-retest

4.1.1. Dotazník osobního štěstí

Tento dotazník (značíme $stesti^{TR}$) vznikl volným překladem z *Oxford Happiness Questionnaire*[15]. Byl vyvinut psychology M. Argylem a P. Hillsem na půdě Oxfordské univerzity a dnes se běžně užívá ve světě experimentální ekonomie či psychologie. Jeho administrace je jednoduchá a test samotný se považuje za standart při měření aktuálního pocitu štěstí. Dotazník byl použit jako odrazový můstek pro praktickou část této práce a zároveň figuruje jako jediný zástupce profesionálně sestaveného a také převzatého dotazníku.

Překlad dotazníku není doslovný, jelikož se při pilotním testování ukázaly některé doslovné překlady jako nevyhovující a špatně srozumitelné. Např. položka č.6 v doslovném překladu zní: „Nejsem úplně optimistický o své budoucnosti“, což se jeví poněkud neohrabaně. Proto bylo tvrzení přeformulováno na „Budoucnost nevidím moc optimisticky“. Překlad položek byl uzpůsoben tak, aby vyzněl co nejpřirozeněji a zapadl do českého prostředí. Původní dotazník je k nahlédnutí pod webovou adresou uvedenou v seznamu literatury.

Dotazník se skládá z 29 položek, které jsou formulovány jako tvrzení, která se týkají dlouhodobějšího rozpoložení a hodnocení života. Respondent ke každému tvrzení vybírá jednu z číslic na škále 1-6, kde 1 reprezentuje úplný nesouhlas a 6 úplný souhlas. 12 z celkových 29 tvrzení je formulováno reverzně, což znamená, že je tvrzení ve tvaru negativní věty. Jako příklad poslouží položka č.5: „Zřídka kdy se probouzím odpočatý“. Zahrnutím takovýchto položek (viz ukázka na obrázku 4.1) se předchází jednak vlivu příliš negativního či pozitivního vyznění položek, jednak nesvědomité vyplňování respondenta (stálé označování jediné možnosti a podobně).

Položky skórujeme pomocí číslic zvolených v rámci jednotlivých odpovědí, tedy $X_{ij} \in \{1, 2, 3, 4, 5, 6\}$. Pro reverzní položky uvažujeme reverzní hodnotící škálu a dle toho i skórujeme, tedy $X_{ij}^{rev} = 7 - X_{ij}$. Pokud respondent v položce dosáhne skóre 1, tak označil odpověď, která značí nejméně pociťované štěstí. Naproti tomu skóre 6 značí pravý opak.

Interpretace celkového skóre je zřejmá - čím vyšší je dosažený počet bodů, tím budeme respondenta považovat za šťastnějšího. Maximálně je v dotazníku možné získat 174 bodů (29 x 6), což interpretujeme jako pocit úplného osobního štěstí. Nejhorší výsledek, kterého je možné dosáhnout je 29 bodů a znamená pravý opak. Jako průměrná hodnota by se dalo očekávat skóre ve výši 101,5 bodů (29 x 3,5).

Respondenti tohoto dotazníku byli studenti 2. a 3. ročníku Obchodní akademie a VOŠ Valašské Meziříčí. První testování proběhlo na konci školního roku v červnu 2016. Retest proběhl krátce po začátku dalšího školního roku v září. Pro zachování anonymity se v dotazníku, který byl v tištěné papírové podobě, nevyžadovalo vyplnění jména a příjmení.

Aby bylo možné po obou fázích dotazníky spárovat, byl každému respondentovi přidělen kód skládající se z jeho iniciál a data narození. Po vyhodnocení první fáze testování se kolonka pro vyplnění tohoto kódu jevila jako slabina, jelikož ji mnoho respondentů nevyplnilo. Po upravení polohy kolonky a jejím zvýraznění se tento problém během retestu neopakoval. Spárovat se ve výsledku povedlo celkem 29 účastníků (12 žen a 17 mužů), což se pro výpočet koeficientu reliability touto metodou jeví jako dostatečné číslo. Zpracování výsledků bylo prováděno ručně, kdy z papírových dotazníků byly převáděny hodnoty do tabulky v počítači, která při výpočtech byla dále zpracovávána softwarem RStudio.

Jako zajímavost uvedeme, že průměrné skóre v testu bylo 118,5 a v retestu 113,9 bodů, což se nám může jevit jako pocit nadprůměrného osobního štěstí. T-test na hladině významnosti 95% zamítl hypotézu o rovnosti středních hodnot, ve prospěch alternativy, že průměrné skóre v retestu je menší než v testu (t-statistika = 1,918; p-value = 0.03268). Znění alternativy bylo formulováno s domněnkou, že žáci se na počátku školního roku cítí méně šťastnými než na jeho konci, což

potvrdil i výsledek.

Podkladem pro výpočet kritériální validity byly dvě dodatečné položky, jejichž zodpovězení nemělo vliv na celkové skóre. Jednalo se o hodnocení toho, jak šťastným se respondent cítí a také to, jak šťastné vnímá respondenty jejich okolí (blízcí přátelé, rodiče). V obou případech mohly být odpovědi na tyto doplňující otázky zkresleny subjektivitou respondenta, což mohlo zkreslit výslednou hodnotu koeficientu kritériální validity.

4. K většině lidí mám vřelý, laskavý vztah.

1 2 3 4 5 6

5. Zřídka se probouzím odpočatý.

1 2 3 4 5 6

6. Budoucnost nevidím moc optimisticky.

1 2 3 4 5 6

Obrázek 4.1: Ukázka položek dotazníku osobního štěstí.

4.1.2. Test matematických dovedností

Jako zástupce výkonového testu pro tuto metodu byl sestaven test matematických dovedností, který budeme značit M^{TR} . V 10 položkách jsou testovány základní dovednosti jako je řešení lineárních rovnic, počítání se zlomky, vyjadřování neznámé proměnné, práce s procenty, zápis intervalů a základy geometrie (viz ukázka 4.2). Úroveň požadovaných znalostí není vyšší než ta, které se dosáhne ve druhém ročníku neodborné střední školy. V každé z položek je v nabídce pět možných odpovědí, přičemž správná je právě jedna. Za každou správně zvolenou odpověď byl respondentovi připsán jeden bod a celkové skóre bylo tvořeno jako součet všech získaných bodů. Položky byly řazeny s převážně narůstající obtížností.

Respondenti byli i v tomto případě studenti 2. a 3. ročníku Obchodní akademie a VOŠ Valašské Meziříčí. Z výsledků testové formy vyhodnocené na konci školního roku v červnu 2016 se počátkem dalšího školního roku v září podařilo spárovat 27

testů. Mezi těmito respondenty bylo zahrnuto 11 žen, 13 mužů a ve 3 odpovědích nebylo pohlaví uvedeno.

Volba termínu testování se jevila příznivě, jelikož jsme v průběhu prázdnin nepředpokládali, že by mohlo dojít k nárůstu pravého skóre (znalostem respondenta) procesem učení. Pravděpodobnější se jevila možnost zapomenutí některých dovedností a znalostí, o čemž svědčil i nižší průměrný počet bodů získaný v retestové fázi. Průměrné skóre v testové formě před prázdninami bylo 6,96 bodů a po prázdninách 6,48 bodů. Nulovou hypotézu o rovnosti středních hodnot, kterou jsme otestovali párovým t-testem proti jednostranné alternativě, že hodnota průměrného skóre v retestové fázi je menší než ve fázi testu, nebylo možné na hladině významnosti 95% zamítnout.

5. Do posloupnosti 4, 6, 10, 18, __, 66 doplň chybějící číslo:

- a. 22
- b. 34
- c. 45
- d. 54
- e. 46

6. Pokud přečte Karolína 15 stran za 20 minut, jak dlouho jí bude trvat přečtení 48 stran?

- a. 60 min
- b. 64 min
- c. 62 min
- d. 58 min
- e. 56 min

Obrázek 4.2: Ukázka položek testu matematických dovedností.

Test byl respondentům předložen v papírové podobě a jeho vyplnění trvalo přibližně 25 minut. Vyhodnocení probíhalo nejprve ruční opravou obdržených odpovědí a následným převedením výsledků do počítačové podoby. Při vyhodnocování se, na rozdíl od dotazníku osobního štěstí, jevíly některé zodpovězené testy jako „sabotážní“, jelikož někteří respondenti naschvál zatrhli všechny položky stejnou odpovědí nebo neodpovídali vůbec, což okomentovali i humornými vzkazy. Takovéto reakce byly spíše výjimkou, ale i tak musely být vyřazeny, aby nezkruslovaly hodnoty prováděných výpočtů. Při vyhodnocení retestových forem tyto

dvě odpovědi navíc nebyly spárovány.

Podkladem pro výpočet kritériální validity toho testu byly známky z matematiky za poslední dvě pololetí, které by měly představovat objektivní kritérium pro validaci.

4.2. Dotazníky sestavené pro metodu paralelních forem

4.2.1. Dotazník projevů nervozity při veřejném vystupování

Tento dotazník (v diskuzi značený jako N^{PF}) byl sestaven dle předpokladů pro výpočet reliability metodou paralelních forem uvedených v odstavci 3.3.2. To znamenalo vytvořit dvě paralelní testové formy s identickou strukturou položek o vlastnostech popsaných výše. V jednotlivých dvojicích paralelních položek se jednalo především o významovou ekvivalenci, shodné počty možností, bodové ohodnocení apod. Jelikož by se vyplňování dvou forem dotazníků s podobně znějícími otázkami mohlo jevit respondentovi podivně a mohlo by vést k ovlivňování odpovědí na jednotlivé položky, byla provedena následující úprava. Veškeré položky byly seskupeny do jednoho dotazníku o celkem 30 položkách, aby se celá forma jevila respondentovi uceleně. V jednotlivých částech testu bylo vždy stejné zastoupení položek spadajících do jednotlivých testových forem, přičemž jejich pořadí bylo neuspořádané. Při vyhodnocení a skórování se test rozdělil zpět na 2 původní formy o 15 položkách a body se do celkového skóre připsaly dle toho, do jaké z paralelních částí položka původně patřila.

Každá z testovaných forem měla tuto strukturu:

- čtveřice položek formulovaná pomocí sémantického diferenciálu zabývající se fyzickými projevy (viz ukázka na obr. 4.3);
- čtveřice položek obsahující tvrzení, které respondent ohodnotil pomocí Likertovy škály (*ano/spíše ano/ spíše ne/ne*);
- šestice položek obsahující tvrzení, které respondent ohodnotil pomocí Li-

kertovy škály (*určitě ano/ano/spíše ano/ spíše ne/ne/určitě ne*);

- jedna situační položka, kde respondent volil možnost, dle které by se zachoval;
- položka pro podklad kritériální validity, která zjišťovala, nakolik nervózního (klidného) shledává respondenta jeho blízké okolí. Ohodnocení této odpovědi přirozeně nemělo vliv na výši celkového skóre.

Každá položka v sobě zahrnovala možnosti, které v sobě nesly informaci o stupni nervozity. Dle toho bylo provedeno také skórování, kdy se slovní odpovědi překódovaly na čísla tak, že možnosti položky s obsaženou informací o vyšším stupni nervozity obdržely vyšší bodové ohodnocení (skóre) a naopak. V souladu s tímto postupem skórování byly obsaženy i reverzní položky, které zamezovaly příliš negativnímu či pozitivnímu vyznění testu jako celku.

1. Představte si, že stojíte před velkou skupinou cizích lidí a máte prezentovat výsledky své práce. Lidé v publiku se mezi sebou baví, musíte se chopit slova, uklidnit je a začít hovořit.

Na dané stupnici zakroužkujte tu možnost, se kterou byste se v rámci této situace nejvíce ztotožnili.

Hovořím klidně a plynule	1	2	3	4	5	Koktám, chvěje se mi hlas
Velice zčervenám	1	2	3	4	5	Vůbec nezčervenám
Nepotím se	1	2	3	4	5	Začne mě polévat pot
Třesou se mi ruce/kolena	1	2	3	4	5	Mám klidný a pevný postoj

Obrázek 4.3: Ukázka dotazníku projevů nervozity

Tento dotazník byl z původní verze sestavené v Microsoft Wordu (součást přílohy) převeden do elektronické podoby přes internetovou službu pro tvorbu dotazníků a průzkumů *Vyplňto.cz*. Respondentům byl v této podobě k dispozici po dobu jednoho měsíce a byl distribuován skrze sociální sítě. Během této doby dotazník vyplnilo 47 respondentů (29 žen a 18 mužů), kteří byli ve věku 16 až 51 let. Díky internetové službě byly po ukončení testování k dispozici exporty

dat v elektronickém formátu, které se po překódování použily pro další zpracování a výpočty pomocí RStudia. Díky elektronickému a zautomatizovanému zpracování bylo značně sníženo riziko dopuštění se chyby při administraci, což je považováno za velké pozitivum. Služba *Vyplňto.cz* poskytla i zajímavé informace z průběhu testování jako například průměrnou délku strávenou vyplňováním dotazníku, která v tomto případě činila 5 minut a 59 vteřin.

4.2.2. Test matematických dovedností

Tento test, který je v diskuzi značený jako M^{PF} , byl vytvořen jako zástupce výkonnostního testu pro metodu paralelních forem. Skládá se z celkem 12 otázek, které jsou tvořeny dvěma podcelky o 6 položkách, které byly v jednotlivých dvojicích paralelní. Tyto položky byly v náhodném pořadí respondentovi opět představeny v jediné formě, protože by se mu zodpovídání dvou velice krátkých testových forem s podobně znějícíma otázkami, mohlo jevit podivně, což by mohlo vést ke zkreslení výsledků. Sjednocením do jedné formy tak bylo docíleno toho, že respondent považoval všechny položky za jeden celek a nebyl ve velké míře znepokojován opakováním se podobností ve spárovaných otázkách. Na seřazení položek dle obtížnosti nebyl kladen velký důraz, jelikož se svou náročností jeví triviálně.

Položky pokrývaly základní matematické dovednosti v rozsahu základní školy. Konkrétně se jednalo o problematiku poměru, zlomků, doplnění členu do číselné posloupnosti, úměry, prvočísel a řešení lineárních rovnic. U položek byla správná vždy jen jedna možnost, za kterou se připisoval jeden bod. Maximálně se tedy dalo získat 6 bodů v každé formě, resp. 12 celkově. Doplnující otázkou sloužící dále jako podklad k výpočtu kriteriální validity byl dotaz, kterou známku respondent nejčastěji dostává (dostával) ve škole z matematiky.

V první fázi tvorby byl test sestaven v Microsoft Wordu (viz příloha) a mezi respondenty byl následně šířen v elektronické podobě pomocí *Vyplňto.cz*. Celkem bylo získáno 50 respondentů - 26 žen a 24 mužů. Jelikož test nebyl zaměřen na konkrétní populaci (jako např. dotazníky test-retest metody vyplněné studenty

SŠ), nebyl vznesen dotaz na věk. Administrace proběhla stejně jako u předchozí metody. Test matematických dovedností byl časově náročnější a průměrná doba vyplnění činila 11 min a 14 vteřin. Je to zapříčiněno tím, že u tohoto testu bylo třeba provést drobné výpočty a zamyšlení se nad správnou odpovědí, kdežto u dotazníku týkajícího se nervozity šlo o prosté vyjádření názoru.

4.3. Dotazníky sestavené pro metodu split half

4.3.1. Dotazník projevu nervozity při veřejném vystupování

Jelikož metoda split half oproti metodám test-retest a paralelním formám neklade na formu testu natolik striktní požadavky, bylo sestavení dotazníku o poznání jednodušší. Abychom mohli testové položky při výpočtu koeficientu reliability touto metodou rozdělit na dvě poloviny, musel být zachován jejich sudý počet.

Tento dotazník, který dále značíme jako N^{SH} , se skládal ze dvou skupin položek. V první skupině bylo obsaženo 20 položek, kde respondenti vybírali možnosti z Likertovy škály (*určitě ano/ano/spíše ano/ spíše ne/ne/určitě ne*), které vyjadřovaly míru jejich ztotožnění se s popsanou situací. Další skupinu tvořila šestice položek ve formě sémantického diferenciálu, kde bylo třeba zvolit stupeň, který respondenti zaujímal mezi dvěma bipolárními hodnotami. Stupeň byl vyjádřen čísly 1-5.

Při vyhodnocení se jednotlivé odpovědi převedli na číselné ohodnocení. Čím většího číselného (a tím i bodového) ohodnocení respondent v položce dosáhl, tím znatelněji jeho odpověď reflektovala zastoupení nervozity. U reverzních položek tomu bylo přesně naopak a během zpracování bylo nutné tyto případy překódovat.

Proces distribuce i administrace byl proveden v elektronické podobě jako u předchozí metody (tj. pomocí služby Vyplňto.cz a RStudia). Celkem bylo získáno 118 respondentů - 72 žen a 46 mužů ve věku 16 až 51 let. Průměrná doba vyplňování byla 4 min a 39 vteřin. Dotazník měl také vysokou návratnost, který činila 75%. To znamená, že při 100 zobrazení dotazníku, jej uživatelé v 75

případech celý vyplnili a odeslali k vyhodnocení.

Jako podklad pro výpočet kritériální validity, byla zvolena otázka na to, jak respondenta vnímá jeho nejbližší okolí (rodina, přátelé).

4.3.2. Test matematických dovedností

Tento test (v diskuzi značen jako M^{SH}) se skládal z 20 položek, které pokrývaly základní matematické dovednosti v rozsahu maximálně 1. ročníku střední školy. Položky se týkaly počítání se zlomky, procenty, jednoduchých slovních úloh, trojčlenky, prvočísla, apod. Každá položka dávala na výběr vždy 5 možných odpovědí, ze kterých byla správná právě jedna. Za každou správnou odpověď byl respondentovi přičítán jeden bod a maximálně tak bylo možno dosáhnout skóre 20 bodů. Aby respondent nebyl ze začátku testu zneklidněn, byly jako první položky voleny ty s nižší obtížností. Obtížnost těch následujících byla různorodá.

Distribuce a následné vyhodnocení bylo identické jako u předchozích elektronicky šířených dotazníků (testů). Získáno bylo 58 respondentů - 30 žen a 28 mužů ve věku 16 - 51 let. Průměrná délka vyplňování byla 20 min a 5 vteřin. Návratnost tohoto testu činila 52%, což je o poznání méně než v případě dotazníků o projevech nervozity. Je to zřejmě způsobeno jednak vyšší časovou náročností pro vyplňování, jednak i obtížností položek, kde se hledá správné tvrzení, nikoliv postoj.

Podkladem pro výpočet kritériální validity posloužila doplňující otázka na známku, kterou respondent nejčastěji získával z matematiky. Ta přirozeně neměla vliv na obdržené celkové skóre testu. Jako zajímavost poznamenejme, že průměrné skóre obdržené respondenty činilo 14,8 bodů. Jelikož se toto skóre jeví nadprůměrně (očekávali bychom jako průměrné skóre 10 bodů), stálo by za zvážení zvýšení obtížnosti jednotlivých položek. Na výpočet koeficientu reliability tato zdánlivě nízká obtížnost testu nemá razantní vliv.

Kapitola 5

Představení výpočtů

V předchozí kapitole jsme si představili dotazníky, které byly sestrojeny pro výpočet koeficientu reliability jednotlivými metodami. Rovněž jsme se seznámili s jejich administrací a skórováním. Z obdržených odpovědí byla v průběhu administrace vytěžena data, která jsme očistili v programu Microsoft Excel. Na těchto datech jsme pak provedli výpočty pomocí programu RStudio.

Na jednotlivých dotaznících si podrobněji představíme výpočet koeficientu reliability vždy tou metodou, pro kterou byl celý dotazník zkonstruován a následně zvážíme, zda-li má smysl vyčíslvat koeficienty reliability i ostatními metodami. Pokud ano, tak to provedeme a v diskuzi níže se zamyslíme, čím mohou být obdržené hodnoty zkresleny.

Pro přehled uvádíme tabulku 5.1, která shrnuje základní charakteristiky a údaje z předchozí kapitoly.

Tabulka 5.1: Srovnání dotazníků a testů

	# respondentů	# položek	distribuce	typ testu
$stesti^{TR}$	29	29	ruční	osobnostní
M^{TR}	27	10	ruční	matematický
N^{PF}	47	30	digitální	osobnostní
M^{PF}	50	12	digitální	matematický
N^{SH}	118	26	digitální	osobnostní
M^{SH}	58	20	digitální	matematický

5.1. Dotazník osobního štěstí ($stesti^{TR}$)

Metoda test-retest

Tento dotazník byl sestaven pro výpočet koeficientu reliability metodou test-retest. Pro lepší představu o povaze dat přikládáme první šestici pozorování (viz obrázek 5.1) a Graf 1, na kterém lze vidět kladnou závislost mezi pozorovanými skóre. Koeficient reliability v případě této metody vypočítáme pomocí vztahu

##	x.score_test	x.score_retest	x.prazdniny	x.nazor_rodicu
## 1	129	126	5	5
## 2	126	116	5	3
## 3	110	141	6	6
## 4	78	63	3	3
## 5	138	125	6	2
## 6	159	143	6	6

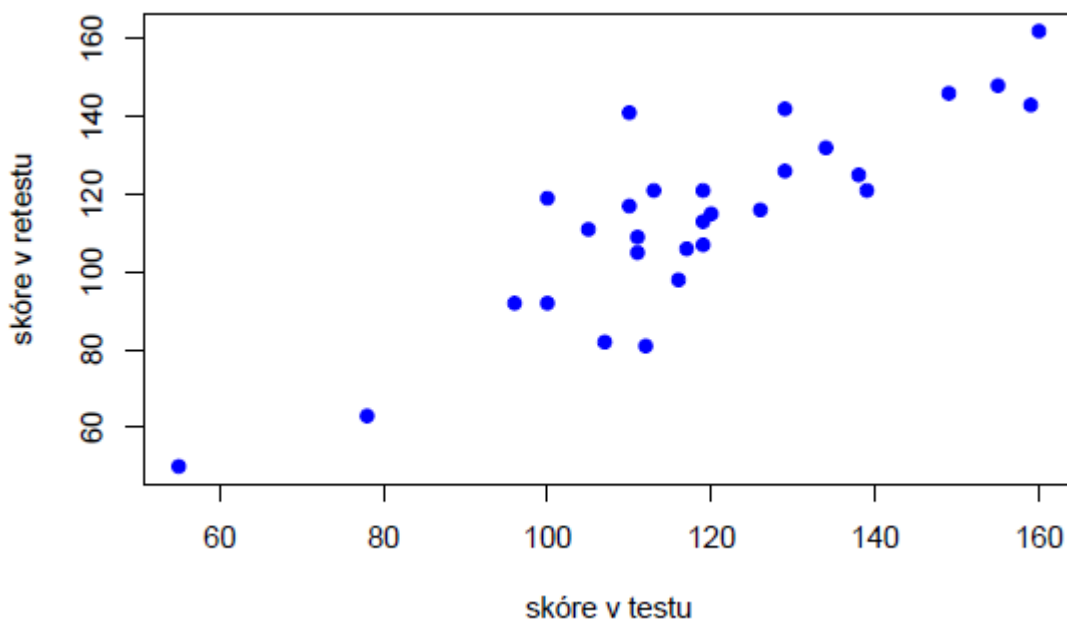
Obrázek 5.1: Ukázka dat z dotazníku osobního štěstí

(2.12) jako Pearsonův korelační koeficient. Při ručním dosazování bychom získali

$$r_{xy}^{TR} = \frac{29 * 405349 - 3436 * 3304}{\sqrt{29 * 421568 - 3436^2} \sqrt{29 * 394304 - 3304^2}} \doteq 0,86. \quad (5.1)$$

Kriteriální validita

Pro podklad kriteriální validity byly v retestové fázi zahrnuty dvě dodatečné položky, které měly stejnou formu jako ostatní položky testu. Tyto položky představovaly dvě validační kritéria, pomoci nichž byly vypočteny dva koeficienty kriteriální validity. Položky se dotazovaly na to, jak respondent hodnotí prožité letní prázdniny a také, jak šťastného vnímají respondenta rodiče a jeho blízké okolí. Hodnotu těchto koeficientů kriteriální validity jsme vyčíslili jako korelaci mezi průměrným skóre za oba testy a hodnotou kritéria. Obdrželi jsme postupně hodnoty 0,65 a 0,52.



Obrázek 5.2: Graf 1

Ostatní metody

Na dotazníku osobního štěstí má smysl dále uvažovat výpočet koeficientu metodami split half a Cronbachovým alfa. Pro metodu split-half je třeba sudý počet položek, kterého dosáhneme vynecháním např. poslední položky (nebo jeho zahrnutím do obou polovin položek). Koeficient reliability spočítáme pomocí korelačního koeficientu (2.12), který dosadíme do Spearmanova-Brownova vzorce (2.13). Získali jsme $r^{SH} \doteq 0,93$.

Výpočet Cronbachova alfa jsme provedli pomocí vzorce (2.15) pro $p = 29$. Pomocí kódu na obrázku 5.3 jsme také zjistili, že $S_X^2 \doteq 516,47$ a $\sum_{j=1}^{29} S_j^2 \doteq 55,00$. Dosazením jsme získali $r_\alpha^R \doteq \frac{29}{28} \left(1 - \frac{55,00}{516,47}\right) \doteq 0,93$. Pro výpočet standardizovaného koeficientu alfa jsme nejprve standardizovali (viz obr. 5.4) hodnoty dle (2.26) a následně dosazovali stejným postupem jako u hrubého koeficientu alfa do vztahu (2.15). Obdrželi jsme hodnotu $r_\alpha^{st} \doteq \frac{29}{28} \left(1 - \frac{27,56}{445,37}\right) \doteq 0,97$.

```
rozpyl_score <- var(x$score_test)
rozptyly_polozek <- c(rep(0,29))
for (i in 1:29){
  rozptyly_polozek[i] <- var(x[,i+1])
}
sum(rozptyly_polozek)
r_alfa_hruby <- (29/28)*(1 - sum(rozptyly_polozek)/var(x$score_test))
```

Obrázek 5.3: Ukázka skriptu

```
for (i in 1:29){
  for (j in 2:30){
    x[i,j] <- ((x[i,j]) - mean(x[,j]))/sd(x[,j])
  }
}
```

Obrázek 5.4: Ukázka skriptu

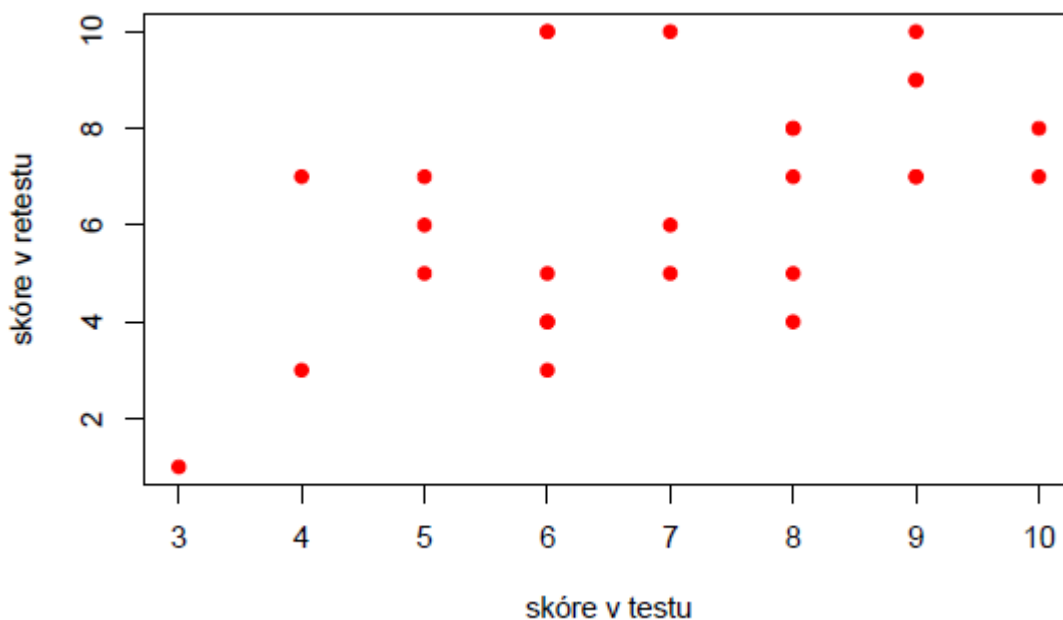
5.2. Test matematických dovedností (M^{TR})

Metoda test-retest

Pro lepší představu o naměřených skóre si můžeme na grafu 2 prohlédnout naměřené hodnoty. Ty působí seskupeně v jednotlivých sloupcích, což je způsobeno celočíselným skórováním, které značí počet správných odpovědí v jednotlivých testech. Koeficient reliability vypočítáme opět jako korelaci a získáváme $r_{xy}^{TR} \doteq 0,51$. retest vypočítáme jako korelaci mezi těmito skóry.

Kriteriální validita

Podkladem pro výpočet koeficientu kriteriální validity byl dotaz na prospěch za poslední dvě pololetí z matematiky. Koeficient v tomto případě vypočítáme jako korelaci mezi průměrným prospěchem za poslední dvě pololetí a průměrným skóre v testu a retestu. Získali jsme hodnotu -0,60, což nás na první pohled může zaskočit, ale uvědomme si její interpretaci v tomto případě. Při větších hodnotách průměrného skóre je tendence dosahovat menších hodnot ve prospěchu a obráceně. Člověk mající vynikající znalosti matematiky by měl dosahovat vysokého skóre testu a zároveň nízké hodnoty prospěchu. Proto je záporná korelace



Obrázek 5.5: Graf 2

na místě a reflektuje velmi dobrou hodnotu kriteriální validity. Pro lepší interpretaci ji v závěrečné tabulce uvádíme s kladným znaménkem.

Ostatní metody

Použití metody split-half na tento dotazník je diskutabilní, jelikož každá polovina testu obsahuje pouze 5 položek, což může vést ke zkreslení výsledku. I přes to jsme pro ilustraci tento výpočet provedli na datech z testové formy rozdělené na sudé a liché položky. Hodnotu korelačního koeficientu jsme dosadili do Spearmanova-Brownova vzorce (2.13) a získali koeficient reliability $r^{SH} \doteq 0,62$.

Mělo smysl uvažovat výpočet koeficientu reliability principem vnitřní konzistence. Jelikož existuje právě jedna správná odpověď u položky, má smysl provést výpočet pomocí Kuderova-Richardsonova vzorce č. 20 (2.21), který je ekvivalencí Cronbachova hrubého alfa pro tento typ položek. Průběh výpočtu, pro který je mj. nutné vyjádřit tzv. difficulty index, je znázorněn v kódu na obrázku 5.6.

Získali jsme $r^{KR} \doteq 0,64$.

```
difficulty_index <- c(rep(0, 10))
for (i in 1:10)
{
  difficulty_index[i] <- mean(x[,i+4])
}
p <- difficulty_index
q <- 1 - p

rozptyl_score <- var(score_liche + score_sude)
sum(p*q)

rel_kr <- (10/(10-1))*(1 - (sum(p*q)/var(score_test)))
```

Obrázek 5.6: Ukázka skriptu

Výpočet standardizovaného alfa jsme nyní provedli užitím vztahu (2.27), který je ekvivalentní pro dosazení standardizovaných skóre do vztahu (2.15). Do vztahu (2.27) dosazujeme průměrnou korelaci, kterou jsme zjistili pomocí vyjádření z korelační matice (viz obr. 5.7). Získali jsme hodnotu $r_{\alpha}^{st} \doteq \frac{10}{9} \left(1 - \frac{5,25}{17,02}\right) \doteq 0,77$.

```
cor_matice <- cor(x[,5:14])
sumaprvek <- 0
for (i in 1:10){
  for (j in 1:10){
    if (i<j)
      sumaprvek <- sumaprvek + cor_matice[i,j]
  }
}
prum_korelace <- sumaprvek/(10*9/2)
stand_cr_alfa <- (10*prum_korelace)/(1+9*prum_korelace)
```

Obrázek 5.7: Ukázka skriptu

5.3. Dotazník projevů nervozity (N^{PF})

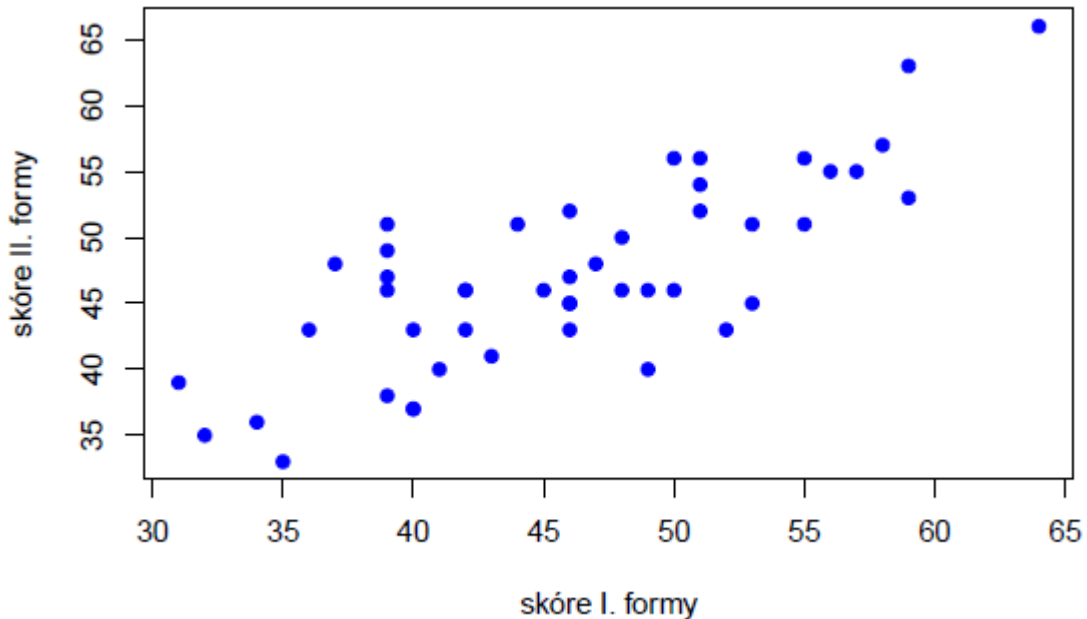
Metoda paralelních forem

Tento dotazník byl sestaven pro metodu paralelních forem a skládal se z 30 položek. Pro názornost si představíme ukázkou z datasetu. Pro zjištění reliability touto metodou byl dotazník vnitřně strukturován do dvou skupin položek

	vek	pohlavi	nazor_okoli	item1	item2	item3	item4	item5	item6	item7
1	16	2	2	3	5	4	3	2	4	3
2	16	2	1	2	4	5	4	3	4	3
3	21	2	2	4	4	3	4	4	3	4
4	21	2	3	1	3	2	4	2	4	3
5	21	1	2	3	2	3	4	3	3	4
6	22	2	3	1	1	2	2	3	2	2

Obrázek 5.8: Ukázka datasetu

splňující požadavky v souladu s kapitolou (2.3.2). Pro výpočet bylo třeba uspořádat položky jednotlivých paralelních forem uspořádat do dvou zvláštních datasetů, ze kterých jsme následně zjistili hodnoty skóre, které jsou vykresleny v Grafu 3. Výpočet koeficientu reliability se provedl pomocí vztahu (2.12), tedy jako výpočet korelačního koeficientu mezi skóre získané v jednotlivých skupinách otázek (viz ukázka 5.10). Získali jsme hodnotu $r^{PAR} \doteq 0,79$.



Obrázek 5.9: Graf 3

```
forma1 <- data.frame(x[,8+c(1,2,3,4,9,11,12,13,14,19,20,21,22,23,24)])
forma2 <- data.frame(x[,8+c(5,6,7,8,10,15,16,17,18,25,26,27,28,29,30)])
r_par <- cor(score1, score2)
```

Obrázek 5.10: Ukázka kódu

Kriteriální validita

Podkladem pro výpočet koeficientu kriteriální validity byla položka, ve které respondent označoval, jak nervózním jej vnímá jeho okolí. Koeficient validity jsme vyjádřili jako korelaci mezi průměrným skóre v jednotlivých formách a kritériem. Výsledná hodnota koeficientu $-0,33$ je záporná, což znamená, že respondenti vyhodnocení testem jako nervózní mají tendenci být svým okolím označování za spíše klidné (a obráceně). Výše tohoto koeficientu reflektuje špatnou validitu testu. Výsledek je však diskutabilní, jelikož položku tohoto kritéria vyplňovali sami respondenti přičemž je zjevné, že respondent sám může jen stěží objektivně vnímat názor okolí na jeho samého. Proto je pravděpodobné, že tato položka podléhala značnému subjektivnímu zkreslení, což by vysvětlovalo i výslednou hodnotu této korelace.

Ostatní metody

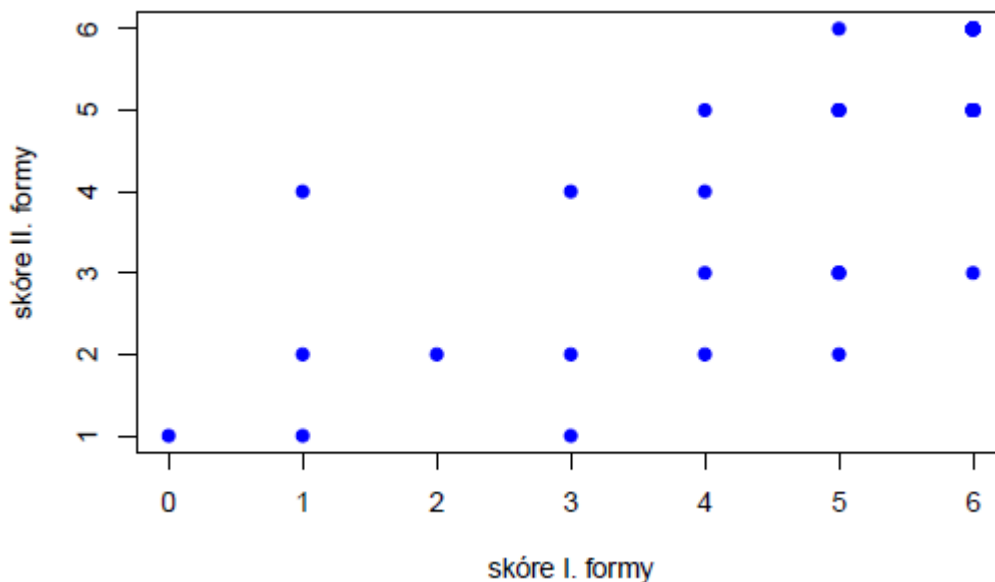
Dále jsme vyjádřili koeficient reliability metodou split-half aplikovanou na sudých a lichých položkách. Pomocí vztahu (2.13) jsme získali $r^{SH} \doteq 0,80$. Vztahem (2.15) jsme vyjádřili koeficient hrubého alfa $r_{\alpha}^R \doteq \frac{30}{29} \left(1 - \frac{41,05}{197,38}\right) \doteq 0,82$. Následným dosazením standardizovaných položek do téhož vztahu jsme vyčíslili koeficient standardizovaného alfa jako $r_{\alpha}^{st} \doteq \frac{30}{29} \left(1 - \frac{28,01}{369,73}\right) \doteq 0,96$.

5.4. Test matematických dovedností (M^{PF})

Metoda paralelních forem

Tento test, blíže popsáný v kapitole 4, se skládal z 12 položek zkoumající úroveň matematických znalostí. Respondentům byl předložen jako jeden celek,

ale jeho vnitřní struktura je v souladu s předpoklady tvořena dvěma paralelními formami o 6 položkách. Koeficient reliability vypočteme jako korelaci mezi celkovými skóre v jednotlivých formách, které jsou zobrazeny v Grafu 4 (viz 5.11. Je zjevná kladná lineární závislost, kterou jsme také vyjádřili korelačním koeficientem, jehož hodnota je zároveň rovna koeficientu reliability $r^{PAR} \doteq 0,78$.



Obrázek 5.11: Ukázka kódu

Kriteriální validita

Podkladem pro výpočet koeficientu kriteriální validity byla položka na prospěch, který respondent nejčastěji získával z matematiky. Stanovíme jej opět jako korelační koeficient, tentokrát mezi hodnotami prospěchu a součtu skóre v jednotlivých formách. Jelikož považujeme za lepší prospěch ve škole ten s menší hodnotou, převrátili jsme u této veličiny znaménko. Obdobný problém nastal i u testu matematických znalostí pro metodu test-retest. Získali jsme hodnotu kriteriální validity 0,57.

Ostatní metody

Metodou split-half použitou na sudých a lichých položkách jsme získali $r^{SH} \doteq 0,89$. Jelikož mají jednotlivé položky testu jednu správnou odpověď, užili jsme pro výpočet hrubého alfa Kuderův-Richardsonův vzorec č.20 (2.28). Obdrželi jsme hodnotu $r^{KR} \doteq 0,87$. Hodnotu standardizovaného alfa jsme získali dosazením standardizovaných hodnot do vztahu (2.15), a sice $r_{\alpha}^{st} = 0,89$.

5.5. Dotazník projevů nervozity (N^{SH})

Tento dotazník se skládal z celkem 26 položek, které se zabývaly projevy nervozity. Byl sestaven pro výpočet metodu split-half, která na jeho sestavení neklade nikterak přísné požadavky. Ukázka obdržených dat je na obrázku 5.12.

	x.vek	x.okoli	item1	item2	item3	item4	item5	item6	item7
1	22	4	2	2	1	2	5	5	5
2	23	2	1	4	1	4	5	4	2
3	18	1	4	5	3	4	4	5	4
4	24	5	2	1	2	3	1	3	2
5	23	2	3	5	3	4	5	6	5
6	20	1	3	3	2	3	3	3	3

Obrázek 5.12: Ukázka kódu

Metoda split-half

V první řadě jsme výpočet touto metodu provedli na sudých a lichých položkách, přičemž jsme obdrželi hodnotu $r_{sude,liche}^{SH} \doteq 0,87$. Pro představu o povaze skóre v těchto polovinách je k dispozici Graf 5. Metodu jsme rovněž vyzkoušeli na náhodně vygenerovaných dvou polovinách testových položek (viz ukázka kódu na obr. 5.14). V tomto případě jsme získali $r_{random}^{SH} \doteq 0,83$. Vidíme, že se hodnoty těchto dvou koeficientů navzdory jiné volbě dvou polovin příliš neliší, což svědčí o vysoké reliabilitě tohoto dotazníku.



Obrázek 5.13: Graf 5

```
x <- nerv_split
set.seed(55)
a <- sample(1:26,13) #generujeme indexy první skupiny položek

vzorek1 <- data.frame(matrix(c(rep(0,118*13)), nrow = 118))
for (i in 1:13){
vzorek1[,i] <- x[ , (a[i]+8)]
}
score1 <- c(rep(0,118))
for (i in 1:118){
score1[i] <- sum(vzorek1[i,])
}
b <- setdiff(1:26,a) #druha skupina otazek je komplementem te první
vzorek2 <- data.frame(matrix(c(rep(0,118*13)), nrow = 118))
for (i in 1:13){
vzorek2[,i] <- x[ , (b[i]+8)]
}
score2 <- c(rep(0,118))
for (i in 1:118){
score2[i] <- sum(vzorek2[i,])
}
r <- cor(score1,score2)
r_sh <- 2*r/(1+r)
```

Obrázek 5.14: Ukázka kódu

Kriteriální validita

Podkladem pro kriteriální validitu byla položka, kde respondent uváděl, jak nervózního jej vnímá jeho okolí. Koeficient kriteriální validity jsme spočítali jako korelaci mezi celkovým skóre v testu a tímto kritériem. Při výpočtech tohoto koeficientu bylo zjištěno opačné (chybné) kódování této otázky, kterou jsme opravili záměnou znaménka. Obdrželi jsme hodnotu 0,45.

Ostatní metody

Stejně jako u předchozích dotazníků jsme vyjádřili hrubý a standardizovaný koeficient alfa. Získali jsme hodnoty $r_{\alpha}^R \doteq 0,82$ a $r_{\alpha}^{st} \doteq 0,96$.

5.6. Test matematických dovedností (M^{SH})

Tento test sestavený pro metodu split-half byl tvořen celkem 20 položkami, které zkoumaly úroveň matematických znalostí. Existovala vždy jedna správná možnost, za kterou byl připisován jeden bod do celkového skóre. Níže v Grafu 6 níže znázorněna skóre v jednotlivých formách testu.

Metoda split-half

Obdobně jako u předchozího dotazníku jsme spočítali koeficienty jednak pro sudé a liché položky, jednak pro náhodně vygenerované dvě poloviny položek. Získali jsme $r_{sude,liche}^{SH} \doteq 0,81$ a $r_{random}^{SH} \doteq 0,89$.

Kriteriální validita

Jako validační kritérium byl položen dotaz na nejčastěji získávaný prospěch v matematice. Kriteriální validitu vypočteme opět jako korelaci mezi kritériem a celkovým skóre v testu. Hodnoty testu jsme dosazovali s opačným znaménkem, abychom vyjádřili, že prospěch s menší hodnotou hodnotíme jako lepší. Hodnota koeficientu byla vypočítána ve výši 0,53.



Obrázek 5.15: Graf 6

Ostatní metody

Jelikož položky měly jednu správnou odpověď, provedli jsme výpočet pomocí Kuderova-Richardsonova vzorce č. 20 (2.28) a obdrželi jsme $r^{KR} \doteq 0,85$. Stejným postupem jako je uváděno výše jsme vypočítali standardizované alfa. Získali jsme $r_{\alpha}^{st} = 0,90$.

Kapitola 6

Diskuze

6.1. Srovnání koeficientů reliability

Nyní si shrneme výsledky, které jsme obdrželi při výpočtech v předešlé kapitole. Ty jsou uvedeny v tabulce 6.1 níže, kde jsou zvýrazněny hodnoty těch koeficientů, pro jejichž výpočet byl daný dotazník/test sestaven. Ostatní (neoznačené) koeficienty reliability byly na daných dotaznících technicky proveditelné, ale vlivem jejich konstrukce či porušení předpokladů mohlo dojít ke zkreslení jejich hodnot. Pro interpretaci hodnot koeficientů reliability se odkazujeme na tabulku 2.4 a obdobně, pro hodnoty kritériální validity, na tabulku 3.3. V obou případech je zřejmé, že hodnota koeficientu reliability/kritériální validity je tím lepší, čím blíže je jedničce.

Tabulka 6.1: Vypočítané hodnoty koeficientů

KOEFICIENT (metoda)	$stesti^{TR}$	M^{TR}	N^{PF}	M^{PF}	N^{SH}	M^{SH}
r_{xy}^{TR} (test-retest)	0,86	0,51				
r^{PAR} (paralelní formy)			0,79	0,78		
$r_{sude,liche}^{SH}$ (split-half)	0,93	0,62	0,80	0,89	0,87	0,81
r_{random}^{SH} (split-half)					0,83	0,89
r_{α}^R resp. r^{KR}	0,93	0,64	0,82	0,87	0,82	0,85
r_{α}^{st}	0,97	0,76	0,96	0,89	0,96	0,90
koef. kritériální validity	0,65; 0,52	0,60	-0,33	0,57	0,45	0,53

Podívejme se nyní na tabulku 6.1 a shrňme si výsledky. Metodou test-retest jsme v případě dotazníku štěstí dosáhli výsledku 0,86, což považujeme za dobré

a svědčící o vysoké reliabilitě tohoto dotazníku. K vysoké napozorované korelaci (reliabilitě) mohla přispět samotná povaha zkoumaného atributu štěstí, vysoký počet položek a také to, jak byly formulované. Týkaly se především dlouhodoběji pocítovaného štěstí, a z tohoto důvodu nepředpokládáme, že by u respondentů během uplynutí více než dvou měsíců mezi testováními došlo k výraznější změně v hodnotě pravého skóre. Na druhou stranu mohlo být ovlivněno termínem retestu, který byl prováděn na počátku školního roku, což mohlo studentům přispívat ke snižování pocitu štěstí. Navzdory tomu obdrženy výsledek referuje o vysoké reliabilitě testu ve smyslu stability v čase.

V případě testu M^{TR} byla vypočítána relativně nízká hodnota 0,51. Domníváme se, že nižší hodnota tohoto koeficientu mohla být způsobena na jedné straně úbytkem matematických znalostí přes prázdniny a na straně druhé, nižším počtem otázek v testové formě, kterých bylo pouze 10. Oproti dotazníku štěstí, kde bylo položek 29, se v obdržených skóre mohla více projevit chybová složka. V případě jejího působení se výsledné skóre ovlivnilo (ať už pozitivně či negativně) v daleko větší proporcii, nežli tomu mohlo být v případě dotazníku štěstí, který má takřka třikrát více položek. Nižší počet položek mohl být tedy vedle prázdnin dalším faktorem snižujícím hodnotu reliability. Naopak lze při srovnávání reliability těchto dvou testů vyloučit vliv počtu respondentů, který byl přibližně stejný.

Metodou paralelních forem byly vypočítány koeficienty reliability na dotazníku N^{PF} a testu M^{PF} . V prvním případě nabyly koeficient reliability hodnoty 0,79 a v druhém 0,78. Obě hodnoty jsou uspokojivé a svědčí o přiměřené reliabilitě těchto testových forem.

Na testových formách N^{SH} a M^{SH} , které byly sestaveny pro metodu split-half, dosáhly koeficienty hodnot 0,87 a 0,81 při seskupení položek na sudé a liché, resp. 0,83 a 0,89 při jejich náhodném rozčlenění. Všechny tyto hodnoty koeficientů klasifikujeme jako dobré a svědčící o vysoké reliabilitě testů, o čemž svědčí i to, že při odlišném výběru položek se hodnota koeficientu příliš nezměnila.

Za povšimnutí stojí hodnoty koeficientů reliability vypočítané metodou split-half na dotaznících/testech sestavených pro metody test-retest a paralelních fo-

rem. Ty jsou vždy vyšší než hodnoty koeficientů vypočítané metodami, pro které byly sestaveny. U dotazníku štěstí a testu M^{TR} je tomu tak, protože jsme výpočet r^{SH} provedli pouze na datech získaných během vyhodnocení 1. testu (nikoliv retestu). Zřejmě má smysl předpokládat vyšší hodnotu korelace mezi skóre v jednotlivých polovinách položek z prvního testu než mezi skóre získanými v testové a retestové formě zvlášť. Tento příklad názorně ukazuje, že metodou split-half neodhadujeme reliabilitu ve smyslu stability výsledků v čase, ale spíše, jako soudružnost a konzistenci jednotlivých polovin testu.

U paralelních forem N^{PF} a M^{PF} lze vyšší hodnotu koeficientu reliability metodou split-half opodstatnit na základě toho, že v původních paralelních formách, které byly spojeny do jediného testu, se po dvojicích spárované položky vyskytovaly v náhodném pořadí. To při jejich následném rozdělení na skupiny sudých a lichých položek při metodě split-half mohlo s velkou pravděpodobností způsobit, že se některé spárované položky ocitly ve stejné skupině a tím ovlivnily (navýšily) výslednou hodnotu koeficientu korelace i reliability.

Dále byly na představené šestici testových forem vyjádřeny koeficienty reliability pomocí Cronbachova hrubého a standardizovaného alfa. V případě testů matematických dovedností, jejichž položky měly vždy jednu správnou odpověď, se pro výpočet hrubého koeficientu alfa užil Kuderův-Richardsonův vzorec, který je pro tento typ položek totožný s původním vzorcem. Z tabulky 6.1 lze vypozařovat, že hodnoty hrubého koeficientu alfa byly vždy obdobné, popř. lehce vyšší, než u hodnot získaných předchozími metodami.

Ze všech koeficientů reliability bylo nejvyšších hodnot dosahováno pomocí Cronbachova standardizovaného alfa. To, na rozdíl od hrubého alfa, které koeficient vyjádří na základě průměrné kovariance mezi položkami a jejich průměrného rozptylu, vyčísľuje koeficient reliability na základě průměrných korelací. Standardizované alfa lze také vyjádřit na základě vztahu pro výpočet hrubého alfa, avšak dosazením standardizovaných položek. Oba tyto postupy jsou ekvivalentní, protože kovariance standardizovaných položek je rovna jejich korelaci a rozptyl položek je normován na jedničku, což lze vidět ve vztahu (2.27).

Při výpočtu obou těchto koeficientů je zohledňována míra asociace odpovědí na jednotlivé položky, nebo též konzistence položek, pomocí velmi podobných charakteristik - kovariance a korelace. Hrubý koeficient alfa je výhodnější používat v případě, kdy chceme při výpočtu reliability zohlednit různé hodnoty rozptylů napříč položkami. Ty mohou být způsobeny jak vlivem různého počtu možností u jednotlivých položek (tzn. formátu), ale také jejich různou obtížností. V případě výkonnostních testů, kdy tato situace nastává, má smysl dávat přednost právě hrubému alfa.

Naproti tomu standardizované alfa má větší smysl při konstantních rozptylech, kdy neuvažujeme různorodost rozptylů jednotlivých položek. Užití standardizovaného alfa v opačném případě je tedy diskutabilní a mohlo by vést k nadhodnocení reliability. Poznamenáváme, že jsme ve zdrojích, ze kterých čerpáme, nenalezli odpověď na to, kterému z koeficientů alfa dát v jaké situaci přednost.

Srovnání dotazníků a testů dle koeficientů reliability

Mohli bychom se ještě pokusit srovnat podle získaných koeficientů reliability námi použité dotazníky/testy. Nejvyšších hodnot napříč metodami dosahuje dotazník osobního štěstí, což lze to vysvětlit profesionálním sestavením, vysokým počtem položek i jejich povahou a formátem. Koeficienty reliability se také jeví vysoce bez ohledu na námi zvolenou metodu. Dotazník se jeví reliabilně jednak ve smyslu časová stability výsledků (metoda test-retest), ale i jako konzistence jednotlivých částí testů.

Naproti tomu nejnižších hodnot koeficientů reliability je dosahováno v testu M^{TR} . V případě metody test-test retest jsme se pokusili nízkou reliabilitu vysvětlit vlivem prázdnin, které mohly ovlivnit hodnotu pravého skóre a tím i stabilitu výsledků. Metodou split-half jsme obdrželi vyšší koeficient než u předchozí metody, ale vlivem nízkého počtu položek lze výsledek považovat za zkreslující. Koeficienty hrubé a standardizované alfa svým přístupem měří konzistenci testu, kterou v tomto testu uvažujeme rovněž nízkou, jelikož každá z položek zkoumala trochu odlišné znalosti z oblasti matematiky (zlomky, rovnice, geometrie,...).

Ostatní dotazníky a testy vykazaly při výpočtech dobré hodnoty reliability v rozmezí přibližně 0,80-0,95. Výsledky považujeme za dobré, ale dle již výše uvedených důvodů uvažujeme s velkou rezervou koeficienty vyjádřené metodou split-half na dotazníku N^{PF} a testu M^{PF} , které byly sestaveny pro metodu paralelních forem.

6.2. Srovnání hodnot kritériální validity

V případě koeficientu kritériální validity jsme až na jeden případ dosáhli velmi dobrých hodnot v rozmezí od 0,45 do 0,65. V dotazníku N^{TR} , který tvořil jedinou výjimku, však vyšla hodnota koeficientu nejen níže, ale dokonce i záporně (-0,33). Výsledek lze na základě dat interpretovat tak, že respondenti vyhodnocovaní testem jako nervóznější se sami považovali za spíše klidné. Tato závislost se jeví zajímavě a možná by stála za důkladnější prozkoumání. Je pozoruhodné, že se tato závislost projevila pouze u dotazníku N^{TR} a v případě N^{PF} navzdory velmi podobné konstrukci testu již nikoliv. Zůstává otázkou, zda příčinou odlišné hodnoty těchto koeficientů je nějaká námi neurčená chyba ve formulaci položek nebo samotná povaha respondentů, kteří v případě těchto dotazníků byli odlišní. Veškeré hodnoty koeficientů kritériální validity i tak uvažujme s jistou rezervou a skepsí, protože na hodnoty validačních kritérií byli dotazováni sami respondenti, což mohlo zapříčinit subjektivitu získaných odpovědí, která mohla vést ke zkreslení výsledných hodnot.

Zjišťování validity jsme se v naší práci věnovali jen okrajově. Pro lepší stanovení kritériální validity by bylo na místě například sestavit další dotazníky, které by měly za úkol objektivněji stanovit hodnotu validačního kritéria. Jako další možnost pro získání lepší hodnoty kritéria by v případě testů nervozity mohl být posudek psychologa a v případě testů matematiky názor pedagoga.

6.3. Shrnutí

Závěrem se ještě vraťme ke koeficientům reliability. Jak vidíme, u každého testu či dotazníku jsme jich obdrželi několik a jejich výši jsme mohli srovnat buď vzájemně nebo ve vztahu k doporučovaným hodnotám. Toto srovnání nám však stále nedává odpověď na to, kterou z metod pro odhad reliability použít. Všechny koeficienty se ve své podstatě snaží odhadnout totéž (reliabilitu), ale k výpočtu přistupují na základě různých hledisek. Metoda test-retest odhaduje reliabilitu srovnáním výsledků z testu předneseném respondentům ve dvou časových momentech a reprezentuje reliabilitu ve smyslu stability výsledků v čase. Metoda paralelních forem přistupuje k reliabilitě opět pomocí podobnosti dvou sad výsledků, ale ty jsou v tomto případě získány v jeden časový moment ze dvou paralelně sestavených testových forem. Metodou split-half při administraci rozčleníme test na dvě poloviny a zkoumáme, nakolik jsou výsledky těchto dvou celků provázány. Pokud zobecníme myšlenku metody split-half, můžeme zkoumat, jak spolu souvisejí nejen poloviny, ale i jednotlivé položky. To se již zabýváme principem vnitřní konzistence testu a výpočty pomocí Cronbachova alfa, které jsou zobecněním metody split-half.

Jak vidíme, každá z uvedených metod přichází s trochu jinou myšlenkou, jak reliabilitu odhadnout a zároveň s sebou nese jistá omezení pro samotnou konstrukci testu. Úplně jasnou a obecnou odpověď na to, kterou z metod pro odhadnutí reliability použít, v praxi nikdy nedostaneme. Je třeba vždy zohlednit situaci, ve které je testování prováděno, a také formu, kterou je dotazník/test sestaven. Další možností je nejprve zvolit způsob, jakým budeme reliabilitu zjišťovat a teprve dle toho sestavit test. Pokud zvolíme tento způsob, budou hlavním aspektem pro volbu metody náklady a pracnost celého sestavení testové formy, která je nejvyšší u metody paralelních forem. Je-li zajištěno splnění veškerých teoretických předpokladů, pak právě tato metoda dává nejkvalitnější výsledky. Úplné splnění předpokladů je však v praxi velice náročné, a proto vždy zvážíme, jakou z metod upřednostníme. Jelikož každá z metod nahlíží na reliabilitu jinak, je různá i interpretace zjištěných koeficientů.

Jak vidíme, výběr vhodné metody je dosti komplexním problémem a musíme do něj zahrnout jak znalost teoretických předpokladů, tak náklady, které jsou úměrné pracnosti sestavování. Je dobré si také uvědomit, že na rozdíl od naší práce je hlavním účelem psychologických testů zkoumání pravého skóre respondenta. Dosažení dobrých hodnot reliability a validity je kritériem nezbytným pro kvalitní testování s relevantními výsledky, ale i tak se jedná spíše o jeho nástroj, nežli o samotný účel.

Závěr

Hlavním cílem této práce bylo seznámit čitatele s koncepty reliability a validity. V první části této práce jsme se věnovali krátkému úvodu do psychometrie a teorii testů, ve které jsme kladli hlavní důraz na model klasické teorie testů, pomocí něhož jsme se seznámili s teoretickým modelem reliability. Ten jsme rozšířili do praktické roviny a uvedli jsme si jednotlivé přístupy, kterými lze míru reliability odhadovat. Dále jsme navázali kapitolou, která představovala problematiku validity a jednotlivé směry, kterými na ni můžeme nahlížet.

Hlavní smysl praktické části spočíval v aplikování teoretických poznatků na problematiku výpočtu reliability. Představili jsme si postupně dotazníky a testy, které jsme pro účely této práce sestrojili a také s průběhem testování a administrací. Na datech získaných vyhodnocením responsí jsme provedli výpočty koeficientů reliability pomocí představených metod a také výpočty zaměřené na vyjádření koeficientu kriteriální validity.

Výpočtům jsme věnovali celou kapitolu, kterou jsme doplnili o několik grafických výstupů a ukázek kódu z programu RStudio, ve kterém byly výpočty prováděny. V závěrečné diskuzi jsme se zamysleli nad obdržnými výsledky, jejich interpretací a možnosti užití jednotlivých metod.

Literatura

- [1] McDonald, Roderick P.: *Test theory: a unified treatment*. Lawrence Erlbaum Associates, Inc., Publishers, New Jersey, 1999
- [2] Chadha, Narender Kumar,: *Applied Psychometry*. SAGE Publications Pvt. Ltd, India, 2009
- [3] Urbánek T., Denglerová D., Širůček J.: *Psychometrika: měření v psychologii*. Praha: Portál, 2011
- [4] Freeman, Frank Samuel: *Theory and Practise of psychological testing*. New York, 1962
- [5] Moosbrugger H., Kelava, A.: *Testtheorie und Fragebogenkonstruktion.*, Springer Medizin Verlag, Heidelberg, 2007
- [6] Trochim William M.K.: *Measurement Validity Types*. In: Web Center for Social Research Methods [online], 2006, [cit. 2017-09-03]. Dostupné z: www.socialresearchmethods.net/kb/measval.php
- [7] Trochim William M.K.: *Reliability and Validity*. In: Web Center for Social Research Methods [online], 2006, [cit. 2017-06-03]. Dostupné z: www.socialresearchmethods.net/kb/relandval.php
- [8] *Understanding Test Quality-Concepts of Reliability and Validity*. In: HR-Guide.com [online], 2015, [cit. 2017-07-03]. Dostupné z: www.hr-guide.com/data/G362.htm
- [9] Jan Schubert: *Klasická testová teorie reliability v metodologii výběrových šetření*. Sociologický ústav AV ČR, Praha, 2010
- [10] Morton S. Mc., Church A.H., Waclawski J.: *Alternative Validation Strategies.*,John Wiley and Sons, Inc., 2007
- [11] Gelman A., Cortina J.: *A Quantitative Tour of the Social Sciences.*, Cambridge University Press, 2009

- [12] Zumbo B. D., Chan E.H.K.: *Validity and Validation in Social, Behavioral, and Health Sciences.*, Springer International Publishing Switzerland, 2014
- [13] *Lee Cronbach*. In: Wikipedia, the free encyclopedia [online], 2017, [cit. 2017-18-03]. Dostupné z: www.en.wikipedia.org/wiki/Lee_Cronbach
- [14] Crocker L., Aigina J.: *Introduction to Classical and Modern Test Theory*, Cengage Learning, 2008
- [15] Stephen Wright: *Take the Oxford Happiness Questionnaire*. In: www.theguardian.com [online], 2014, cit. [2016-10-06]. Dostupné z: www.theguardian.com/lifeandstyle/2014/nov/03/take-the-oxford-happiness-questionnaire

PŘÍLOHA: Dotazníky

Dotazník osobního štěstí (test-retest)

--	--	--	--	--	--	--	--

Následuje řada výroků týkajících se lidského štěstí. Označte vždy **jedno z čísel**, které nejlépe odpovídá Vaší situaci dle následující stupnice:

1 = úplně nesouhlasím
6 = úplně souhlasím

Pro zachování anonymity a dobrou práci při vyhodnocení, **vyplňte osmimístný kód v pravém horním rohu** dle následujícího vzoru:

KAREL NOVÁK narozený **01.06.1998** vyplní:

K	N	0	1	0	6	9	8
---	---	---	---	---	---	---	---

Jste: muž x žena

1. Necítím se být úplně spokojený/spokojená se svou situací.

1 2 3 4 5 6

2. Velmi se zajímám o lidi ve svém okolí.

1 2 3 4 5 6

3. Cítím, že můj život je naplňující.

1 2 3 4 5 6

4. K většině lidí mám vřelý, laskavý vztah.

1 2 3 4 5 6

5. Zřídka se probouzím odpočatý.

1 2 3 4 5 6

6. Budoucnost nevidím moc optimisticky.

1 2 3 4 5 6

7. Většina věcí mi přijde zábavná.

1 2 3 4 5 6

8. Jsem vždy oddaný(á) a angažovaný(á) v nějaké činnosti.

1 2 3 4 5 6

9. Život je fajn.

1 2 3 4 5 6

10. Nemyslím si, že by tento svět byl dobrým místem.

1 2 3 4 5 6

11. Hodně se směji.

1 2 3 4 5 6

12. Jsem spokojený se vším v mém životě.

1 2 3 4 5 6

13. Nemyslím si, že jsem pro ostatní atraktivní.
1 2 3 4 5 6
14. Je rozdíl v tom, co bych si přál dělat a tím, co dělám.
1 2 3 4 5 6
15. Jsem velice šťastný.
1 2 3 4 5 6
16. V některých věcech nalézám krásu.
1 2 3 4 5 6
17. Vždy přenáším radost na ostatní.
1 2 3 4 5 6
18. Pro to, co chci dělat, si čas vždy najdu.
1 2 3 4 5 6
19. Mám pocit, že nad svým životem nemám úplnou kontrolu.
1 2 3 4 5 6
20. Mám pocit, že se dokážu postavit čemukoliv.
1 2 3 4 5 6
21. Cítím, že jsem velmi pohotový.
1 2 3 4 5 6
22. Často prožívám radost a nadšení.
1 2 3 4 5 6
23. Dělat rozhodnutí mi dělává problémy.
1 2 3 4 5 6
24. Moc nechápu smysl svého života.
1 2 3 4 5 6
25. Mám energie na rozdávání.
1 2 3 4 5 6
26. Většinou mám věci pod kontrolou.
1 2 3 4 5 6
27. Být s ostatními mě moc nebaví.
1 2 3 4 5 6
28. Necítím se být úplně zdravým.
1 2 3 4 5 6
29. Na minulost nemám moc dobré vzpomínky.
1 2 3 4 5 6
30. Mé prázdniny byly vynikající.
1 2 3 4 5 6
31. Rodiče a přátelé mě považují za šťastného.
1 2 3 4 5 6

Test matematických dovedností (test-retest)

V tomto testu Vás čeká **10 úloh** z matematiky, přičemž každá má **jen jedno správné řešení**. Vyplněním mi pomůžete získat data pro bakalářskou práci, která se týká teorie testů. Účelem není hodnotit, jak je kdo dobrý v matematice, ale spíše, jak dobře je sestaven tento test.

Pro zachování anonymity a dobrou práci při vyhodnocení, **vyplňte individuální osmimístný kód** dle následujícího vzoru:

KAREL NOVÁK narozený **01. 06. 1998** vyplní:

K	N	0	1	0	6	9	8
---	---	---	---	---	---	---	---

• **Váš kód:**

--	--	--	--	--	--	--	--

• **Vaše pohlaví:** **MUŽ x ŽENA**

• **Vaše známky z matematiky v uplynulém roce:**

I. pololetí	II. pololetí

1. Je dán trojúhelník A s délkami stran 12 cm, 8 cm a 16 cm. Trojúhelník B je vzhledem k trojúhelníku A v poměru $\frac{1}{4}$. Která z následujících možností reprezentuje strany takto vzniklého trojúhelníku B?

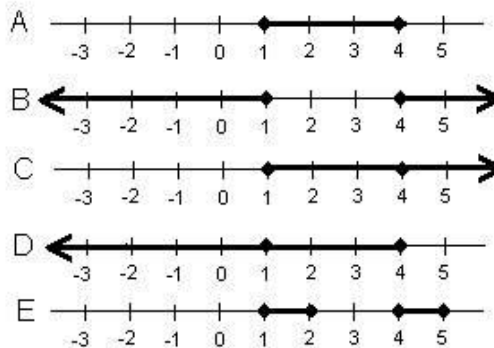
- a. 4 cm, 2 cm, 8 cm
- b. 2 cm, 3 cm, 8 cm
- c. 3 cm, 2 cm, 4 cm
- d. 6 cm, 4 cm, 8 cm
- e. 4 cm, 3 cm, 5 cm

2. Který útvar má 5 stěn a 8 hran?

- a. trojboký jehlan
- b. krychle
- c. pravidelný čtyřstěn
- d. trojboký hranol
- e. pětistěn

3. Která z následujících možností zobrazuje podmínku $x \geq 1 \cap x \leq 4$?

- a. řádek A
- b. řádek B
- c. řádek C
- d. řádek D
- e. řádek E



4. Který z následujících zlomků je menší než $\frac{2}{3}$?

- a. $\frac{7}{8}$
- b. $\frac{5}{6}$
- c. $\frac{3}{4}$
- d. $\frac{3}{5}$
- e. $\frac{5}{7}$

5. Do posloupnosti 4, 6, 10, 18, ____, 66 doplň chybějící číslo:

- a. 22
- b. 34
- c. 45
- d. 54
- e. 46

6. Pokud přečte Karolína 15 stran za 20 minut, jak dlouho jí bude trvat přečtení 48 stran?

- a. 60 min
- b. 64 min
- c. 62 min
- d. 58 min
- e. 56 min

7. Jaké je nejbližší prvočíslo následující po číslu 67?

- a. 68
- b. 69
- c. 71
- d. 73
- e. 76

8. Pokud je úřednice schopna napsat stranu dokumentu za p minut, jakou část strany napíše za 5 minut?

- a. $5/p$
- b. $p-5$
- c. $p+5$
- d. $p/5$
- e. $1-p+5$

9. Pokud $8x + 5x + 2x + 4x = 114$, pak $5x + 3 =$

- a. 12
- b. 25
- c. 33
- d. 47
- e. 86

10. Která z následujících posloupností čísel je uspořádaná od nejmenšího k největšímu?

- a. $-\frac{29}{7}; -\frac{3}{7}; -\frac{52}{26}; 75\%; \frac{5}{6}; \frac{12}{10}$
- b. $-\frac{7}{29}; -\frac{7}{3}; -\frac{52}{26}; 75\%; \frac{5}{6}; \frac{120}{40}$
- c. $-\frac{29}{7}; -\frac{7}{3}; -\frac{52}{26}; 75\%; \frac{5}{6}; 1.2$
- d. $\frac{10}{9}; \frac{9}{8}; \frac{8}{7}; \frac{7}{6}; \frac{6}{7}$
- e. $\frac{7}{6}; \frac{6}{7}; \frac{8}{7}; \frac{9}{8}; \frac{12}{9}$

Dotazník projevů nervozity (paralelní formy)

Dostává se Vám do rukou dotazník zabývající se projevy nervozity při vystupování na veřejnosti. Skládá se ze dvou velice podobných částí a jeho vyplnění Vám nezabere více než 15 minut. Data získaná tímto testem poslouží jako podkladový materiál k bakalářské práci týkající se teorie testů. Děkuji předem za Váš čas a upřímnost obětovanou při vyplňování tohoto dotazníku.

Pohlaví: MUŽ x ŽENA

Věk:

Rodiče a spolužáci (pracovní kolegové) mne považují za člověka:

- a) Zcela klidného a vyrovnaného, který bývá nervózní zřídka
- b) Spíše klidného s občasnými projevy nervozity
- c) Ani klidného, ani nervózního
- d) Spíše nervózního
- e) Velice nervózního člověka, který se nechá lehce

Část I.

1. Představte si, že stojíte před velkou skupinou cizích lidí a máte prezentovat výsledky své práce. Lidé v publiku se mezi sebou baví, musíte se chopit slova, uklidnit je a začít hovořit.

Na dané stupnici zakroužkujte tu možnost, se kterou byste se v rámci této situace nejvíce ztotožnili.

Hovořím klidně a plynule	1	2	3	4	5	Koktám, chvěje se mi hlas
Velice zčervenám	1	2	3	4	5	Vůbec nezčervenám
Nepotím se	1	2	3	4	5	Začne mě polévat pot
Třesu se mi ruce/kolena	1	2	3	4	5	Mám klidný a pevný postoj

2. V této tabulce následuje několik tvrzení týkající se nervozity při veřejném vystupování. Zakřížkujte variantu, která je Vám nejbližší.

	ANO	SPÍŠE ANO	SPÍŠE NE	NE
Vyhýbám se komunikaci s lidmi na přepážce				
Často a rád poznávám nové lidi				
Rád se ujmu vedení při práci ve skupině				
Jsem v úzkých, když se na mě soustředí skupina lidí				

3. V tabulce je řada výroků. Zakřížkujte variantu, se kterou se nejvíce ztotožňujete.

	URČITĚ ANO	ANO	SPÍŠE ANO	SPÍŠE NE	NE	URČITĚ NE
Moc se mi líbí, ale neoslovím ho (ji), bojím se odmítnutí						
Nebudu mu teď volat, určitě nemá čas.						
Snažím se mít věci lepší než ostatní, jinak ztrácím sebevědomí.						
Než abych šel na veřejné WC to zkusím nějak vydržet						
Na poradě neprojevim svůj názor, přestože vím, že mám čím přispět						
Líbilo by se mi být hercem/herečkou						

4. Při obědu ve školní jídelně/menze

- a) Si sedám vždy na stejné místo
 - b) Nezáleží mi na tom, kde sedím
 - c) Sednu si raději osamoceně
 - d) Přisednu si k ostatním
-

Část II.

1. Představte si, že stojíte před velkou skupinou cizích lidí a máte prezentovat výsledky své práce. Lidé v publiku se mezi sebou baví, musíte se chopit slova, uklidnit je a začít hovořit.

Na dané stupnici zakroužkujte tu možnost, se kterou byste se v rámci této situace nejvíce ztotožnili.

Jsem uvolněný	1	2	3	4	5	Mám svíravý pocit, nevolnost
Rychle mi buší srdce	1	2	3	4	5	Mám klidný puls
Situace neovlivňuje můj tělesný komfort	1	2	3	4	5	Pocítuji návaly horka/zimy
Těkám pohledem a nesoustředím se	1	2	3	4	5	Dívám se klidně a zpříma

2. V této tabulce následuje několik tvrzení týkající se nervozity při veřejném vystupování. Zakřížkujte variantu, která je Vám nejbližší.

	ANO	SPÍŠE ANO	SPÍŠE NE	NE
Vadí mi být pozorován ostatními				
Než abych se hádal, ustoupím				
Mám rád, když mě hodnotí ostatní				
Nerad telefonuji				

3. V tabulce je řada výroků. Zakřížkujte variantu, se kterou se nejvíce ztotožňujete.

	URČITĚ ANO	ANO	SPÍŠE ANO	SPÍŠE NE	NE	URČITĚ NE
Při bloudění v neznámém městě se na cestu zeptám hned první osoby						
Raději napíšu e-mail, než abych telefonoval na bezplatnou linku						
Po požití alkoholu snáze navazuji kontakty a jsem k lidem družnější						
Často svůj výkon v práci (škole) srovnávám s ostatními						
Noc před důležitou událostí mám problém s usnutím/spánkem						
Při zkoušce si nemůžu vzpomenout na věci, o kterých vím, že je umím						

4. Po náročném víkendů mě čeká dlouhá cesta vlakem. V kupé je poslední volné místo, na kterém má někdo kufr a já:

- raději zůstanu v uličce
- zeptám se, zdali je místo volné a pak si přisednu
- vejdu dovnitř a posadím se
- zeptám se až ve chvíli, kdy mě začnou bolet nohy

Test matematických dovedností (paralelní formy)

Dostává se Vám do rukou test matematických dovedností. Skládá se 12 úloh mající vždy jedno správné řešení a jeho vyplnění Vám nezabere více než 20 minut. Data získaná tímto testem poslouží jako podkladový materiál k bakalářské práci týkající se teorie testů. Účelem testu není posuzovat, jak je kdo dobrý v matematice, ale spíše jak dobře je sestaven tento test. Děkuji předem za Váš čas a snahu obětovanou při vyplňování.

Pohlaví: MUŽ x ŽENA

Věk:

Známka, kterou jsem nejčastěji dostával z matematiky: 1 2 3 4 5

1. Je dán trojúhelník A s délkami stran 12 cm, 8 cm a 16 cm. Trojúhelník B je vzhledem k trojúhelníku A v poměru $\frac{1}{4}$. Která z následujících možností reprezentuje strany takto vzniklého trojúhelníku B?

- f. 4 cm, 2 cm, 8 cm
- g. 2 cm, 3 cm, 8 cm
- h. 3 cm, 2 cm, 4 cm
- i. 6 cm, 4 cm, 8 cm
- j. 4 cm, 3 cm, 5 cm

2. Který z následujících zlomků je menší než $\frac{2}{3}$?

- f. $\frac{7}{8}$
- g. $\frac{5}{6}$
- h. $\frac{3}{4}$
- i. $\frac{3}{5}$
- j. $\frac{5}{7}$

3. Do posloupnosti 4, 6, 10, 18, ____, 66 doplň chybějící číslo:

- f. 22
- g. 34
- h. 45
- i. 54
- j. 46

4. Pokud přečte Karolína 15 stran za 20 minut, jak dlouho jí bude trvat přečtení 48 stran?

- f. 60 min
- g. 64 min
- h. 62 min
- i. 58 min
- j. 56 min

5. Jaké je nejbližší prvočíslo následující po číslu 67?

- f. 68
- g. 69
- h. 71
- i. 73
- j. 76

6. Pokud $8x + 5x + 2x + 4x = 114$, pak $5x + 3 =$

- f. 12
- g. 25
- h. 33
- i. 47
- j. 86

7. Obdélník má strany délek 6 cm a 8 cm. Zvětšíme-li obě strany v poměru 3:2, jaký bude obsah nově vzniklého obdélníku?

- a) 72 cm^2
- b) 112 cm^2
- c) 108 cm^2
- d) 48 cm^2

8. Který z následujících zlomků je větší než $5/6$?

- a) $2/3$
- b) $6/7$
- c) $20/24$
- d) $49/59$

9. Platí, že $4y + 5y - 7y + 3 = 15$. Jaká je hodnota výrazu $4y - 7$?

- a) 9
- b) 5
- c) 17
- d) 15

10. 4. Které z následujících čísel není prvočíslo?

- a) 83
- b) 101
- c) 87
- d) 29
- e) 53

11. Doplň správné číslo do posloupnosti: 7, 10, 16, 28, ____, 100

- a) 46
- b) 48
- c) 50
- d) 52
- e) 54

12. Jede-li Petr na kole průměrnou rychlostí 20 km/hod, trvá mu cesta do školy 45 minut. Jak dlouho mu bude cesta trvat, pojedě-li průměrnou rychlostí 25 km/hod?

- a) 40 min
- b) 36 min
- c) 35 min
- d) 25 min
- e) 28 min

Dotazník projevů nervozity (split-half)

Dostává se Vám do rukou dotazník zabývající se projevy nervozity při vystupování na veřejnosti, který Vám nezabere více než 10 minut. Data získaná tímto testem poslouží jako podkladový materiál k bakalářské práci týkající se teorie testů. Děkuji předem za Váš čas a upřímnost obětovanou při vyplňování tohoto dotazníku.

Pohlaví: MUŽ x ŽENA

Věk:

Rodiče a spolužáci (pracovní kolegové) mne považují za člověka:

- a) Zcela klidného a vyrovnaného, který bývá nervózní zřídka
- b) Spíše klidného s občasnými projevy nervozity
- c) Ani klidného, ani nervózního
- d) Spíše nervózního
- e) Velice nervózního člověka, který se nechá lehce vykolejit

Následuje několik výroků a situací. Zakřížkujte na řádce vždy tu z variant, která nejvíce vystihuje vaši situaci a je vám nejbližší.

	URČITĚ ANO	ANO	SPÍŠE ANO	SPÍŠE NE	NE	URČITĚ NE
Často svůj výkon v práci (škole, práci ve skupině) srovnávám s ostatními						
Noc před důležitou událostí mám problém s usnutím/spánkem						
Po požití alkoholu snáze navazuji kontakty a jsem k lidem družnější						
Při zkoušce si nemůžu vzpomenout na věci, o kterých vím, že je umím						
Raději napíšu e-mail, než abych telefonoval na bezplatnou infolinku						
Při bloudění v neznámém městě se na cestu zeptám hned první osoby						
Moc se mi líbí, ale neoslovím ho (ji), bojím se odmítnutí						
Nebudu mu (ji) teď volat, určitě na mě nemá čas.						
Snažím se mít věci lepší než ostatní, jinak ztrácím sebevědomí.						

Než abych šel na veřejné WC, to zkusím nějak vydržet						
Na poradě neprojeví svůj názor, přestože vím, že mám čím přispět						
Při obědvání v jídelně mi nedělá problém přisednout si k cizím lidem						
Vadí mi, když je má činnost pozorována ostatními						
Než abych se hádal a vysvětloval svou pravdu, ustoupím						
Mám rád, když mé vystupování hodnotí druzí lidé						
Vyhýbám se komunikaci s lidmi na přepážce (úřady, banky, pošta,...)						
Často a rád se seznamuji s novými lidmi						
Rád se ujmu vedení při práci ve skupině						
Přestože je v kupé jen poslední volné místo, jdu dovnitř a přisednu si						
Nevím, kolik je hodin. Bez váhání se zeptám člověka, co stojí poblíž						

Představte si, že stojíte před velkou skupinou cizích lidí a máte prezentovat výsledky své práce. Lidé v publiku se mezi sebou baví, musíte se chopit slova, ztišit je a začít hovořit.

Na dané škále vyberte vždy tu variantu, se kterou byste se v rámci výše popsané situace nejvíce ztotožnili.

Hovořím klidně a plynule	1	2	3	4	5	Koktám, chvěje se mi hlas
Velice zčervenám	1	2	3	4	5	Vůbec nezčervenám
Vůbec se nezpotím	1	2	3	4	5	Začne mě polévat pot
Jsem uvolněný	1	2	3	4	5	Mám svíravý pocit, nevolnost
Srdce mi buší rychle	1	2	3	4	5	Mám klidný puls
Bez ohledu na situaci se cítím komfortně	1	2	3	4	5	Pocítuji návaly horka/zimy

Test matematických dovedností (split-half)

V tomto testu Vás čeká 20 úloh z matematiky, přičemž každá má **jen jedno správné řešení**. Vyplněním mi pomůžete získat data pro bakalářskou práci, která se týká teorie testů. Účelem není hodnotit, jak je kdo dobrý v matematice, ale spíše, jak dobře je sestaven tento test. Vyplnění by Vám nemělo zabrat více jak 20 minut.

Pohlaví: MUŽ x ŽENA

Věk:

Známka, kterou jste nejčastěji dostávali v matematice: 1 2 3 4 5

1. Je dán trojúhelník A s délkami stran 12 cm, 8 cm a 16 cm. Trojúhelník B je vzhledem k trojúhelníku A v poměru $\frac{1}{4}$. Která z následujících možností reprezentuje strany takto vzniklého trojúhelníku B?

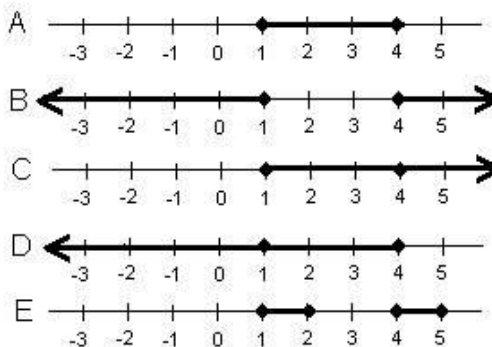
- a) 4 cm, 2 cm, 8 cm
- b) 2 cm, 3 cm, 8 cm
- c) 3 cm, 2 cm, 4 cm
- d) 6 cm, 4 cm, 8 cm
- e) 4 cm, 3 cm, 5 cm

2. Který útvar má 5 stěn a 8 hran?

- a) trojboký jehlan
- b) krychle
- c) pravidelný čtyřstěn
- d) trojboký hranol
- e) pětistěn

3. Která z následujících možností zobrazuje podmínku $x \geq 1 \cap x \leq 4$?

- a) řádek A
- b) řádek B
- c) řádek C
- d) řádek D
- e) řádek E



4. Který z následujících zlomků je menší než $\frac{2}{3}$?

- a) $\frac{7}{8}$
- b) $\frac{5}{6}$
- c) $\frac{3}{4}$
- d) $\frac{3}{5}$
- e) $\frac{5}{7}$

5. Do posloupnosti 4, 6, 10, 18, ____, 66 doplň chybějící číslo:

- a) 22
- b) 34
- c) 45
- d) 54
- e) 46

6. Pokud přečte Karolína 15 stran za 20 minut, jak dlouho jí bude trvat přečtení 48 stran?

- a) 60 min
- b) 64 min
- c) 62 min
- d) 58 min
- e) 56 min

7. Jaké je nejbližší prvočíslo následující po číslu 67?

- a) 68
- b) 69
- c) 71
- d) 73
- e) 76

8. Pokud je úřednice schopna napsat stranu dokumentu za p minut, jakou část strany napíše za 5 minut?

- a) $5/p$
- b) $p - 5$
- c) $p + 5$
- d) $p/5$
- e) $1 - p + 5$

9. Pokud $8x + 5x + 2x + 4x = 114$, pak $5x + 3 =$

- a) 12
- b) 25
- c) 33
- d) 47
- e) 86

10. Která z následujících posloupností čísel je uspořádaná od nejmenšího k největšímu?

- a) $-\frac{29}{7}; -\frac{3}{7}; -\frac{52}{26}; 75\%; \frac{5}{6}; \frac{12}{10}$
- b) $-\frac{7}{29}; -\frac{7}{3}; -\frac{52}{26}; 75\%; \frac{5}{6}; \frac{120}{40}$
- c) $-\frac{29}{7}; -\frac{7}{3}; -\frac{52}{26}; 75\%; \frac{5}{6}; 1.2$
- d) $\frac{10}{9}; \frac{9}{8}; \frac{8}{7}; \frac{7}{6}; \frac{6}{7}$
- e) $\frac{7}{6}; \frac{6}{7}; \frac{8}{7}; \frac{9}{8}; \frac{12}{9}$

11. Tričko stálo původně 138 Kč a bylo zlevněno o 25%. Jaká je jeho cena po slevě?

- a) 67 Kč
- b) 103.50 Kč
- c) 134.50 Kč
- d) 113 Kč
- e) 125 Kč

12. Jaký je kořen následující rovnice: $2y/3 = 8 + 4y$

- a) -2.4
- b) 2.4
- c) 1.3
- d) -1.3
- e) 0

13. Jaký je aritmetický průměr následujících čísel? 5, 12, 6, -2, 4

- a) 3
- b) 5
- c) 4
- d) 6
- e) 2

14. Pokud $y(x-1) = z$, čemu je rovno x ?

- a) $y-z$
- b) $z/y + 1$
- c) $y(z-1)$
- d) $z(y-1)$
- e) $1-zy$

15. Dva úhly v trojúhelníku měří 15° a 85° . Jak velký je třetí úhel?

- a) 50°
- b) 55°
- c) 60°
- d) 80°
- e) 90°

16. Podstavou 7 cm vysokého válce je kruh o poloměru 2 cm. Jaký je jeho objem?

- a) 14π
- b) 28π
- c) 7π
- d) 4π
- e) Žádná z předchozích možností

17. Který z následujících zlomků je roven $\frac{5}{6}$?

- a) $\frac{20}{30}$
- b) $\frac{15}{24}$
- c) $\frac{25}{30}$
- d) $\frac{40}{54}$
- e) $\frac{2}{7}$

18. Kolik mililitrů je v jednom litru?

- a) 10 000
- b) 1 000
- c) 0.1
- d) 0.01
- e) 0.001

19. Kolik je $\frac{1}{9}$ z 9?

- a) $\frac{1}{9}$
- b) 0
- c) 1
- d) 2
- e) 81

20. Odvěsny v pravoúhlém trojúhelníku jsou rovny 3 cm a 4 cm. Jak dlouhá je přepona?

- a) 3 cm
- b) 3,5 cm
- c) 4 cm
- d) 5 cm
- e) 6 cm