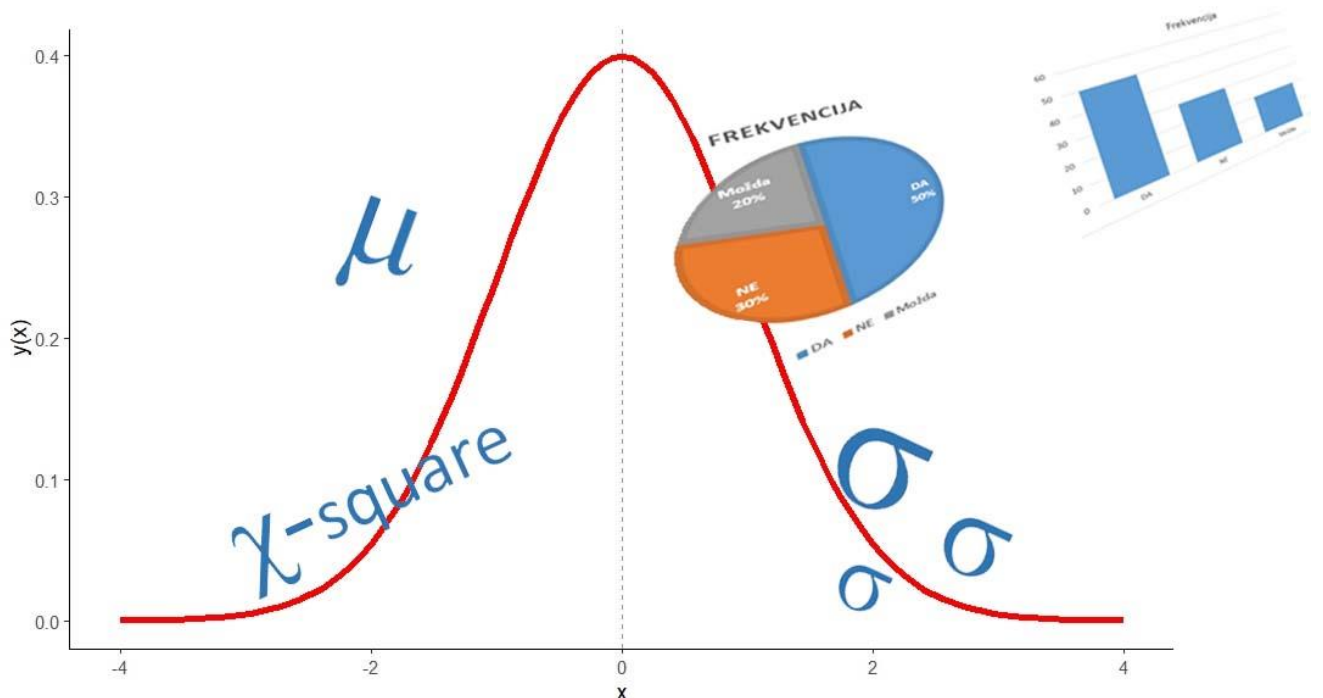




Merima Muslić, Dušanka Bošković, Dženana Husremović

Metode evaluacije korisničkog iskustva za softvere u obrazovanju



Univerzitet u Sarajevu



**Metode evaluacije korisničkog iskustva za
softvere u obrazovanju**

Merima Muslić, Dušanka Bošković, Dženana Husremović

Sarajevo, 2024.

Autori: Merima Muslić, Dušanka Bošković, Dženana Husremović
Naziv djela: Metode evaluacije korisničkog iskustva za softvere u obrazovanju
Izdavač: Univerzitet u Sarajevu – Filozofski fakultet
Za izdavača: prof.dr. Kenan Šljivo

Recenzenti: Prof.dr. Emina Dedić Bukvić
Prof.dr. Selma Rizvić
Prof.dr. Vladimir Hedrih

Ilustracije: Dušanka Bošković
DTP: Dušanka Bošković

Godina izdanja i godina štampanja: 2024.

Odlukom Senata Univerziteta u Sarajevu broj 01-6-31/24 od 27.11.2024. godine data je saglasnost da se udžbenik *Metode evaluacije korisničkog iskustva za softvere u obrazovanju* izda kao univerzitetsko izdanje.

ISBN 978-9926-491-35-2

CIP zapis dostupan u COBISS sistemu Nacionalne i univerzitetske biblioteke BiH pod
ID brojem **62385414**

Sadržaj

Popis slika.....	v
Popis tabela.....	vii
Uvod	1
1. Šta je upotrebljivost i zašto testiramo upotrebljivost softvera.....	5
1.1. Istraživanja upotrebljivosti	6
2. Odabir ispitanika u istraživanju.....	11
3. Mjerenje i mjerni instrumenti	15
3.1. Mjerenje.....	15
3.2. Greške u mjerenju.....	16
3.3. Karakteristike mjernih instrumenata	17
3.3.1. Valjanost.....	17
3.3.2. Pouzdanost.....	18
3.3.3. Diskriminativnost	18
3.3.4. Objektivnost	18
3.4. Reprezentativnost uzorka kod mjerenja	19
3.4.1. Utvrđivanje ciljne populacije	19
3.4.2. Utvrđivanje okvira uzorkovanja	19
3.4.3. Izbor metode uzorkovanja	20
3.4.4. Određivanje potrebne veličine uzorka.....	24
3.5. Upitnici za mjerenje upotrebljivosti	24
3.5.1. Dizajniranje anketnog upitnika.....	26
3.5.2. Izvori greške vezani za anketni upitnik	29
3.6. Standardizirani upitnici za mjerenje upotrebljivosti.....	32
3.7. Ostali načini mjerenja u istraživanjima upotrebljivosti.....	34
3.7.1. Vrijeme završetka zadatka.....	34
3.7.2. Kombinirana mjerenja	34
3.7.3. A/B testiranje.....	34
3.8. Zaključak	35
4. Eksperimentalni nacrti studija upotrebljivosti.....	37
4.1. Eksperimentalni nacrti sa nezavisnim grupama	38
4.1.1. Odabir ispitanika za eksperiment i raspored po grupama	39
4.2. Vrste grupnih eksperimentalnih nacrta.....	40
4.2.1. Nacrt sa kontrolnom grupom i opažanjem prije i poslije intervencije	40
4.2.2. Nacrt sa kontrolnom grupom i mjerenjem samo poslije intervencije.....	41
4.3. Eksperimentalni nacrti sa ponovljenim mjerenjima	41
4.4. Složeni eksperimentalni nacrti	43

4.5. Analiza varijanse za obradu rezultata eksperimentalnih nacrti	44
4.6. Odabir eksperimentalnog nacrti.....	45
4.6.1. Odabir statističkog postupka za obradu podataka	45
4.7. Artefakt u eksperimentu	46
4.8. Kvaziekperiment	48
5. Ne-eksperimentalne metode evaluacije upotrebljivosti.....	51
5.1. Posmatranje/opažanje	53
5.1.1. Vrste posmatranja	54
5.1.2. Vrste ponašanja koje je moguće oponašati.....	57
5.1.3. Analiza podataka prikupljenih posmatranjem	57
5.1.4. Objektivnost mjerenja	58
5.2. Korelacijsko istraživanje	58
5.2.1. Psihologijsko testiranje.....	59
5.3. Tumačenje nalaza korelacijskog istraživanja	59
5.4. Anketno istraživanje.....	60
5.4.4. Načini unapređenja valjanosti anketnih istraživanja	60
5.4.5. Snaga i jasnoća stava	61
6. Pregled statističkih postupaka koje često koristimo	63
6.1. Vrste varijabli	63
6.2. Frekvencije	63
6.3. Mjere centralne tendencije i mjere varijabiliteta	66
6.3.1. Mjere centralne tendencije	66
6.3.2. Mjere varijabiliteta	68
6.4. Raspodjele podataka.....	69
6.4.1. Histogrami	71
6.4.2. Normalizacija podataka.....	75
6.5. Kompariranje dvije aritmetičke sredine (t-test)	77
6.5.1. T-test za zavisne uzorke	77
6.5.2. T-test za nezavisne uzorke.....	78
6.5.3. Interpretacija rezultata T-testa.....	79
6.6. Hi-kvadrat test	80
6.7. Korelacija	81
6.7.1. Koeficijent korelacije	82
6.8. Faktorska analiza	82
7. Modeli prihvatanja tehnologije.....	85
7.1. Model prihvatanja tehnologije - TAM	86
7.2. Objedinjeni teorijski model prihvaćanja i korištenja tehnologije - UTAUT.....	89

8. Teorija kognitivnog opterećenja i upotrebljivost	93
8.1. Teorija kognitivnog opterećenja.....	94
8.1.1. Organizacija pamćenja	95
8.2. Vrste kognitivnog opterećenja.....	96
8.2.1. Intrinzično kognitivno opterećenje.....	97
8.2.2. Irelevantno kognitivno opterećenje	98
8.2.3. Relevantno kognitivno opterećenje	99
8.3. Postupci manipulacije kognitivnim opterećenjem.....	99
8.3.1. Postupci za manipulaciju intrinzičnim opterećenjem.....	99
8.3.2. Postupci za manipulaciju irelevantnim opterećenjem	100
8.3.3. Postupci manipulacije relevantnim opterećenjem	100
8.4. Teorija kognitivnog opterećenja i upotrebljivost: preklapanja.....	103
8.5. Teorija kognitivnog opterećenja i softveri u obrazovanju.....	104
9. Značaj eksplorativne analize prikupljenih podataka	107
9.1. Primjeri vizualizacije kod eksplorativne analize podataka	108
9.2. Primjeri vizualizacije kod komunikacije rezultata	110
Zaključak	113
Literatura	115

Popis slika

Slika 1.1. Fokus različitih disciplina na određene aspekte procesa istraživanja.....	9
Slika 1.2. Algoritam odlučivanja o tipu studije	10
Slika 3.1. Primjer linearne nzmeričke skale ocjenjivanja	27
Slika 3.2. Primjer Likertove skale slaganja	27
Slika 3.3. Primjer matrice ocjenjivanja više tvrdnji Likertovom skalom slaganja	28
Slika 3.4. Primjer skale poređenja	29
Slika 3.5. Primjer SUS upitnika od deset pitanja.....	33
Slika 3.6. Algoritam odlučivanja o studiji evaluacije upotrebljivosti.....	35
Slika 6.1 Primjeri vizualizacije frekvencija korištenjem (a) pita i (b) stupičastog dijagrama	64
Slika 6.2. Primjer vizualizacije frekvencija odgovora na Likertovoj skali korištenjem dijagrama divergentnih stubaca.....	66
Slika 6.3. Primjer funkcije gustoće vjerovatnoće za bacanje jedne kocke	69
Slika 6.4. Primjer distribucije vrijednosti bacanja jedne kocke za eksperiment od 100 bacanja.....	70
Slika 6.5. Primjer funkcije gustoće vjerovatnoće za bacanje tri kocke	70
Slika 6.6. Primjer histograma koji odgovara normalnoj distribuciji.....	71
Slika 6.7. Kriva normalne distribucije	72
Slika 6.8. Uniformna distribucija.....	73
Slika 6.9. Poissonova distribucija	74
Slika 6.10. Eksponencijalna distribucija.....	75
Slika 6.11. Krive normalne raspodjele različite spljoštenosti i iskrivljenosti.....	76
Slika 6.12. Iskrivljena distribucija vremena izvršenja i distribucija vrijednosti dobijenih logaritamskom transformacijom.....	76
Slika 6.13. Dijagram raspršenja kojim se uvezuju visina očeva i sinova	81
Slika 7.1. Model A-TAM.....	87
Slika 7.2. Model TAM-2.....	88
Slika 7.3. Proširenje uticaja na Percipiranu lakoću korištenja kod modela TAM-3.....	89
Slika 7.4. Model UTAUT	90
Slika 8.1. Osnovna kognitivna arhitektura	96
Slika 9.1. Šematske prikaz raspodjele vrijednosti (box-plot) i prikaz specifičnih tačaka na box-plotu i njihova značenja	107
Slika 9.2. Prvi grafikon šematske raspodjele vrijednosti - stupac opsega (Spear, 1952)	108
Slika 9.3. Poređenje grupa podataka pomoću dva box plota	109
Slika 9.4. Grafovi Anscombeovog kvarteta.....	109

Slika 9.5. Tačkasti grafovi za veličine x i y koji ukazuju na njihovu korelaciju	
prije i nakon uklanjanja strčecih vrijednosti (Bošković sur., 2024)	110
Slika 9.6. Divergentni stupičasti grafikoni za ilustraciju odgovora na Likertovoj skali od 5 tačaka : (a) profesori i (b) studenti (Rizvić i sur., 2019)	112

Popis tabela

Tabela 3.1. Pregled skala, empirijskih operacija i primjenjivih statistika po Stivensu	16
Tabela 4.1. Algoritam odlučivanja kod poređenja dvije grupe.....	46
Tabela 4.2. Algoritam odlučivanja kod poređenja tri i više grupa.....	46
Tabela 5.1. Algoritam odlučivanja kod studija sa jednom grupom	53
Tabela 6.1. Primjer frekvencijske tabele.....	64
Tabela 6.1. Primjer frekvencijske tabele iz realnog istraživanja	65
Tabela 6.3. Skala korelacije	82
Tabela 8.1. Efekti kognitivnog opterećenja fokusirani na smanjenje irelevantnog kognitivnog opterećenja.....	101
Tabela 9.1. Primjer upitnik za istraživanje multimedijalnih softvera u nastavi (Rizvić i sur., 2019)	110-111

Uvod

Svi imamo omiljeni model mobilnog telefona, omiljeni operativni sistem, aplikaciju za bankarstvo, društvenu mrežu i sl. Tipični korisnik ne razmišlja sistematično o razlozima preferiranja jednog softvera ili aplikacije u odnosu na drugu. Obično se razlozi u svakodnevnom govoru formuliraju kroz generička objašnjenja da je nešto „lakše za koristiti“, „ljepše“ ili „funkcionalnije“. Softver inženjeri moraju imati detaljnije i sistematičnije informacije o tome šta čini proizvod koji prave privlačnijim u odnosu na druge. Stručan i analitičan pristup traženja odgovora na ovo pitanje se naziva testiranjem ili evaluacijom upotrebljivosti softvera.

Barnum (2010) navodi da u situacijama kada je upotrebljivost inherentna proizvodu koji koristimo, nismo ni svjesni toga da proizvod koji koristimo ustvari vrlo upotrebljiv tj. upotrebljivost je „nevidljiva“. Razlog za ovu pojavu je u tome što nam odgovaraju proizvodi koji imaju ugrađenu upotrebljivost. Ne moramo se mi prilagođavati proizvodu da bismo ga uspješno koristili, proizvod radi onako kako mi želimo tj. kako nam je logično i prirodno. Možda smo morali naučiti nekoliko stvari, ili više od nekoliko, da bismo krenuli, ali ne smeta nam jer je napor bio mali, a nagrade velike. Nagrade o kojima u ovom kontekstu govorimo su lakoća savladavanja funkcija, jednostavnost upotrebe, intuitivnost i zabava.

U situacijama kada smo prinuđeni da koristimo proizvod koji nije upotrebljiv, trošimo puno vremena u savladavanje funkcija te smanjeno nam je povjerenje da koristeći taj softver možemo na ispravan način riješiti problem koji imamo. U tim slučajevima onda pribjegavamo upotrebi dodatnih uređaja ili programa kako bismo bili sigurni u ispravnost i uspješnost urađenog. Navedeno nam troši dodatno vrijeme i energiju. Prvom prilikom kada nađemo adekvatnu zamjensku aplikaciju, softver ili uređaj prestat ćemo koristiti onu koja nam je trošila vrijeme i zaboraviti na njeno postojanje.

Ako ste softver inženjer, web dizajner, vizuelni ili grafički dizajner, stručnjak za korisničku pomoć, odgovorni za podučavanje za upotrebu proizvoda informacione tehnologije (IT) ili neko drugi ko ima učešća u razvoju ili podršci IT proizvoda bilo koje vrste, tada ste upravo vi neko ko treba da razmišlja i zagovara razmišljanje o upotrebljivosti i razvoju visoko upotrebljivog proizvoda. Vaša strast prema korisniku, zagovaranje olakšavanja svakodnevnih aktivnosti korisniku u njegovo ime mogu i utječu na upotrebljivost proizvoda.

Uprkos vašoj želji da podržite potrebe korisnika, možda još uvijek ne vršite testiranje upotrebljivosti ili niste sigurni da to činite na ispravan način. Možda imate problem sa interpretiranjem rezultata koje ste dobili i prevođenjem informacija o upotrebljivosti u upute kako unaprijediti proizvoda ili imate potreba za formaliziranjem i/ili standardiziranjem postupka testiranja upotrebljivosti. Cilj ove knjige je ponuditi neophodne informacije i alate kao odgovor na potrebe istraživača iz područja upotrebljivosti.

U našoj analizi potreba i ciljeva istraživanja upotrebljivosti razlučili smo postojanje dva pristupa istraživanju upotrebljivosti softvera. Prvi ima za cilj dolazak do saznanja koja će biti generalna, saznanja koja će služiti kao svojevrсна uputstva za dizajn novih softverskih

rješenja, novih tehnologija interakcije. Lazar i sur. (2010) ciljeve ove vrste istraživanja klasificiraju na sličan način: razumijevanje novih problema, razumijevanja konteksta upotrebe tehnologije, opis konteksta upotrebe tehnologije i opis upotrebe novih tehnologija. Ova istraživanja zbog tako postavljenih ciljeva moraju imati sofisticiranije razrađenu metodologiju (odabir uzorka, instrumentarij i sl.) kao i primjenu naprednijih statističkih metoda.

Drugi pristup karakterizira fokus na rješenje konkretnog problema u razvoju konkretnog softvera. Cilj ovih istraživanja je na efikasniji način doći do odgovora koja bi intervencija na već postojećoj aplikaciji napravila odgovarajuće poboljšanje, pri tome naravno, uzimajući u obzir i troškove. Ovakva istraživanja, s obzirom na to da nemaju za cilj generalizirati rezultate, imaju u pravilu jednostavniji istraživački nacrt, manji uzorak, instrumentarij dizajniran da dâ odgovore na konkretna pitanja vezano za upotrebu konkretne aplikacije u određenim uslovima od strane ciljne populacije.

Ovaj rukopis ima ambiciju da bude koristan u oba prethodno opisana slučaja istraživanja upotrebljivosti i korisničkog iskustva. Istraživači iz područja upotrebljivosti koji imaju za cilj generalizaciju svojih rezultata i dolazak do novih spoznaja u ovom području će naći korisnim poglavlja u kojima su opisani metodološki i statistički postupci, područja primjene, prednosti i nedostaci određenih pristupa. Značaj kojeg korisnik dobija u procesu razvoja proizvoda se ogleda i u uvođenju novog koncepta korisničkog iskustva (eng. *User eXperience*) – UX koji nadilazi upotrebljivost i obuhvata sveukupnu subjektivnu percepciju korištenja i ispunjenosti očekivanja. Korisničko iskustvo je oblast istraživanja sa fokusom na čovjeka i njegov odnos prema tehnologiji koju koristi, tako da uspješno kombinuje tehnologiju i psihologiju. Donald Norman se smatra jednim od rodonačelnika oblasti interakcije čovjek računar i on je uveo termin korisničko iskustvo u svojoj knjizi „Dizajn svakodnevnih stvari“ (Norman 1988). Interesantno je napomenuti da je originalni naslov knjige bio „Psihologija svakodnevnih stvari“, i da je naslov promijenjen nakon što su autora informisali da se knjiga loše prodaje jer je klasificirana u djela iz psihologije, kako naslov sugerise. Promjena naslova je značajno doprinijela i prodaji i prepoznavanju knjige u krugovima koji se bave dizajnom različitih proizvoda, a posebno softverskih proizvoda. Mislimo da studenti psihologije i psiholozi mogu otkriti da su istraživanja upotrebljivosti i korisničkog iskustva područje kojem mogu doprinijeti svojim kompetencijama i ovom knjigom želimo ukazati na prostor njihovog mogućeg zajedničkog djelovanja sa inženjerima.

Inženjeri koje zanima kako unaprijediti konkretan alat koji razvijaju će naći dosta postupaka i objašnjenja na koji način postaviti istraživanje i posebno, na koji način interpretirati dobijene rezultate kako bi se proizvod unaprijedio. Svako poglavlje će sadržavati osnovno teorijsko objašnjenje relevantnih konceptata i pristupa, kao i praktične primjere iz područja evaluiranja upotrebljivosti i korisničkog iskustva.

U prvom dijelu knjige se daju teoretski osnovi evaluacije upotrebljivosti, tako da se u prvom poglavlju uvode koncepti upotrebljivosti i korisničkog iskustva, zatim motivacije zašto testiramo upotrebljivost softvera, sa posebnim naglaskom na edukacione softvere. U drugom poglavlju su opisani postupci odabira ispitanika u istraživanju upotrebljivosti i značaj ovih postupaka za uspjeh čitavog procesa evaluacije. Treće poglavlje daje pregled pristupa

mjerenju, mjernih instrumenata, skala mjerenja, i uvodi se upitnik kao najčešće korišteni instrument mjerenja upotrebljivosti sa uključenjem korisnika. Kako su ovo najčešće korištene studije u praksi, posebna pažnja je posvećena reprezentativnosti uzorka nad kojim se vrše mjerenja i dizajniranju upitnika i tipičnim vrstama pitanja. U ovom poglavlju se daje i okvirni algoritam odlučivanja o tipu studije koja se primjenjuje za različite kontekste i ciljeve istraživanja, što predstavlja ujedno i pregled tema u drugom dijelu knjige.

Drugi dio knjige je posvećen dizajniranju studija evaluacije upotrebljivosti i korisničkog iskustva. Četvrto poglavlje opisuje dizajniranje eksperimentalnih nacrti istraživanja, dok se peto bavi studijama sa jednom grupom koje se u literaturi označavaju kao ne-eksperimentalne studije. U šestom poglavlju je dat pregled statističkih postupaka koji se koriste u prethodno opisanim studijama.

Treći dio knjige se bavi izdvojenim temama, koje su značajne za upotrebljivost softvera koji se koriste u obrazovanju. Sedmo poglavlje daje pregled najznačajnijih modela prihvatanja tehnologije, koji su određeni percepcijom o korisnosti tehnologije u obavljanju zadatka i percipiranom lakoćom upotrebe. Kada su u pitanju softveri koji se koriste u obrazovanju korisnost je vezana za uspješnost u postizanju određenih ishoda učenja efektivno i efikasno, a lakoća upotrebe je vezana za zadovoljstvo korisnika, odnosno njegovo sveobuhvatno korisničko iskustvo. Osmo poglavlje se bavi teorijom kognitivnog opterećenja u kontekstu upotrebljivosti softvera posebno u obrazovanju. Za dizajn uspješnog edukativnog softvera je ključno razumijevanje kognitivnog opterećenja, koje predstavlja mentalni napor potreban za obradu informacija i izvršavanje zadataka, da bi se izbjeglo da edukativni softver unosi nepotrebno dodatno kognitivno opterećenje. Završno, deveto poglavlje je posvećeno eksplorativnoj analizi podataka koji su prikupljeni tokom istraživanja, kao značajnom, a često zanemarenom koraku u obradi podataka i tumačenju rezultata studija evaluacije upotrebljivosti.

1. Šta je upotrebljivost i zašto testiramo upotrebljivost softvera

Internacionalna organizacija za standardizaciju upotrebljivost definira kao (2018): „Stepen u kojem proizvod može da se koristi od strane određenog korisnika kako bi se efikasno, efektivno i sa osjećajem zadovoljstva postigao određeni cilj relevantan za kontekst upotrebe“. Barnum (2010) u prethodno navedenoj definiciji naglašava važnost tri ključna elementa: (1) određeni korisnik – dakle ne bilo koji korisnik nego korisnik kojem je proizvod namjenjen, (2) određeni cilj – korisnici kojima je namjenjen proizvod imaju istovjetan cilj prilikom korištenja koji je u suštini i cilj postojanja proizvoda, (3) kontekst upotrebe – proizvod mora biti dizajniran u kontekstu u kojem će ga ciljani korisnici koristiti.

Ostali elementi koje naglašava ova definicija su: efikasnost, efektivnost i korisničko zadovoljstvo, kao ključni elementi prilikom mjerenja upotrebljivosti. Kod razmišljanja o upotrebljivosti softvera, tražimo odgovor na pitanje da li korisnik korištenjem ovog alata namjeravanu aktivnost uspijeva završiti na način koji je bolji od trenutno dostupnih. Ovdje ustvari govorimo o dodatnoj vrijednosti koju proizvod daje prilikom izvršenja zadatka na način da npr. smanjuje potrebno vrijeme za izvršenje zadatka ili korisnik lakše izvršava zadatak. Nielsen & Norman Group umjesto definicije upotrebljivost nude opis pomoću slijedećih pet karakteristika: (1) lako se uči; (2) omogućava da se efektivno, efikasno i na zadovoljavajući način obave zadaci; (3) primjenjuje korisniku poznate uzorke dizajna, metafore i modele; (4) minimizira mogućnost greške i korisniku obezbjeđuje mehanizme oporavka; (5) omogućava ugodno korištenje (Nielsen, 2012).

Potrebno je naglasiti važnost korisničkog zadovoljstva prilikom procjene upotrebljivosti. Korisničko zadovoljstvo o kojem se ovdje govori je percipirano zadovoljstvo od strane korisnika. Percipirano zadovoljstvo korištenjem proizvoda znači poželjan proizvod.

Šira potreba za istraživanjem upotrebljivosti softvera javila se u momentu kada su računari ušli u domaćinstva. Do 70-tih godina prošlog stoljeća jedino su profesionalci (inženjeri, matemaričari i sl) koristili računare. Nakon toga, u 80-tim godinama razvija se kućno računalo koje dolazi u domaćinstva slično kao i drugi kućanski aparati. U tom momentu su računare počele koristiti i osobe koje nemaju relevantno obrazovanje, te je sa tim rođena i potreba za prilagođavanjem softvera njihovim iskustvima i očekivanjima, za šta je bilo neophodno evaluiranje - mjerenje njihovog iskustva.

Važno je istaći i da se istovremeno počinju javljati i problemi u industriji gdje se računari uvode u procese upravljanja. Problemi su bili uzrokovani greškama u interakciji operatera i računara. Iako su prve reakcije bile da se ove greške svrstaju u ljudski faktor, izdovjili su se eksperti koji su prepoznali da je zapravo neadekvatan interfejs uzrokom incidenata, i da bi većina operatera postupila isto. Primjer je incident u Three Mile Island (TMI) nuklearnoj elektrani u SAD-u. Jedan od ključnih problema je bio nemogućnost posade,

koja je inače bila vrlo iskusna, da razumije šta se dešavalo u kritičnim momentima u elektrani. To je ukazalo da je sistem nadzora elektrane bio adekvatan za normalan radni režim, ali ne i za situacije kada dođe do havarije (Rogovin i Frampton, 1979). Učešće u ekspertskom timu koji je istraživao ovaj događaj je motivisalo Donalda Normana da svoju djelatnost preusmjeri od obučavanja ljudi kako da izbjegnu greške prilikom korištenja računara ka dizajnu interakcije i principima razvoja upotrebljivih softvera i računarskih sistema (Norman, 1988).

Istraživanja upotrebljivosti je važan dio multidisciplinarnog područja istraživanja interakcija čovjek-računar (eng. *Human-Computer Interaction* – HCI). Istraživanje upotrebljivosti koristi saznanja: računarskih znanosti, psihologije, sociologije, komunikologije, mašinstva i mnogih drugih. U području istraživanja korisničkog iskustva koristimo saznanja i metode ovih znanstvenih područja i modificiramo ih kako bi došli do odgovora na pitanja koja su relevantna u objašnjavanju interakcije čovjek - računar.

Ovakva pozicija istraživačkog područja upotrebljivosti nosi svoje izazove, posebno prilikom evaluacije kvalitete istraživanja. Pored toga, istraživanje upotrebljivosti zahtijeva precizno definirane metode koje su i znanstveno relevantne. Teškoća leži u činjenici da je istovremeno neophodno osigurati znanstvenu relevantnost i praktičnu vrijednost tj. jasne praktične implikacije rezultata za unapređenje softvera. Često je lakše fokus staviti na jednu ili drugu stranu ove vage. Generalno u znanosti imamo područja istraživanja koja za fokus imaju teorijska saznanja, više nego na praktičnu relevantnost. Međutim, istraživanje upotrebljivosti mora biti praktično i relevantno za korisnike, organizacije i dizajnere. Istraživanje mora biti u stanju utjecati na dizajn softvera, procese razvoja proizvoda, obuku korisnika, javnu politiku ili nešto drugo. Upravo u ovom zahtjevu za finim balansom između teorijskog i praktičnog i leži najveći izazov prilikom dizajniranja istraživanja upotrebljivosti. Sva istraživanja upotrebljivosti trebala bi barem uzeti u obzir potrebe i istraživača za identificiranjem znanstvenih činjenica i praktičara kojima su potrebne konkretne primjenjive smjernice. U isto vrijeme, korištene istraživačke metode (bez obzira na izvornu disciplinu) moraju biti rigorozne i primjerene. Nije dovoljno razviti novi računarski interfejs bez istraživanja potrebe za interfejsom i bez praćenja korisničkih procjena tog interfejsa. Istraživači iz područja upotrebljivosti su često stavljeni u poziciju da moraju izaći i uvjeriti druge u potrebu fokusiranja na čovjeka i njegove potrebe i karakteristike u dizajniranju softvera. Jedini način da se potkrijepe izjave o važnosti korisnika i dizajna usmjerenog na čovjeka su čvrsta, rigorozna istraživanja.

Zbog ovog interdisciplinarnog fokusa i historijskog razvoja ove oblasti, postoji mnogo različitih pristupa mjerenju i istraživanju koji se trenutno koriste u oblasti upotrebljivosti. I sami istraživači koji rade na temama vezanim za upotrebljivost se često ne slažu oko toga šta znači „pravo istraživanje upotrebljivosti“. Postoje velike razlike u tome kako različiti lideri u ovoj oblasti vide postojanje interakcije čovjek - računar.

1.1. Istraživanja upotrebljivosti

Lazar i sur. (2017) navode sedam doprinosa istraživanju iz oblasti interakcije čovjek - računar:

1. Empirijska istraživanja— tj. podaci dobijeni empirijskim postupcima (kvalitativni ili kvantitativni) prikupljeni bilo kojom od metoda: eksperimentalni dizajn, ankete, fokus grupe, vremenski dnevnici, senzori i druga automatizirana sredstva, etnografija i dr.
2. Analiza artefakata—dizajn i razvoj novih artefakata, uključujući interfejse, alate, i prototipove. Ovi artefakti su često praćeni empirijskim podacima o povratnim informacijama ili upotrebi. Ova vrsta doprinosa je često poznata kao istraživanje HCI sistema, tehnike interakcije HCI ili prototip HCI dizajna.
3. Metodološki razvoj—novi pristupi koji utiču na procese u istraživanju ili praksi, kao što su nova metoda, nova primjena metode, modifikacija metode ili nova metrika ili instrument za mjerenje.
4. Teorijske analize—koncepti i modeli koji potiču na inovaciju, koji mogu biti prediktivni ili deskriptivni, kao što je teoretski okvir, prostor dizajna ili konceptualni model.
5. Doprinos u vidu objavljivanja skupova podataka—podaci prikupljeni tokom istraživanja, koji daju informacije korisne istraživačkoj zajednici, uključujući repozitorije znanja, benchmarke i skupove podatka .
6. Pregledna istraživanja - pregled i sinteza dostupnih istraživanja u određenoj oblasti, sa ciljem identifikacije trendova i specifičnih tema na kojima je potrebno više rada. Ova vrsta doprinosa može se javiti tek nakon što istraživanja u određenoj oblasti postoje nekoliko godina, tako da ima dovoljno materijala za meta-analizu.
7. Doprinosi raspravi —analize koje nastoje da ubijede čitaoce da se predomisle tj. da prihvate iznijetu argumentaciju, često koriste dijelove drugih prethodno navedenih istraživanja kako bi se određeni stav podržao.

Većina istraživanja interakcije čovjek - računar je empirijskog tipa. Lazar i sur. (2017) navode da se radi o čak 70% istraživanja. Cilj ove knjige je objasniti najčešće korištene postupke u empirijskim istraživanjima te ponuditi primjere za svaki od objašnjenih postupaka i povezati ih sa odgovarajućim formalnim metodama istraživanja, uključujući cjelokupno planiranje i organizaciju istraživanja, odnosno prikupljanje podataka.

U ranim danima istraživanja interakcije čovjek - računar mjerenje se zasnivalo na postupcima za mjerenje postignuća koje nam je ponudila psihologija. Tražio se odgovor na pitanja o tome koliko brzo bi neko mogao da izvrši zadatak ili koliko je zadataka uspješno obavljeno, a koliko grešaka je napravljeno.

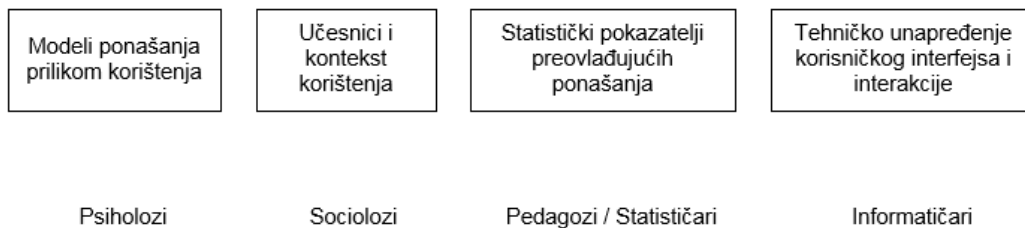
Ovo su i dalje osnovni temelji za mjerenje upotrebljivosti softvera i kao takvi su i danas relevantni. Mjerenja ove vrste su u velikoj mjeri zasnovana na modelu usmjerenom na zadatke, gdje se specifični zadaci mogu odvojiti, kvantificirati i izmjeriti. Pri tome mjerimo tačnost izvršenog zadatka, vremenske performanse, stopu grešaka, vrijeme za učenje, zadržavanje tokom vremena i zadovoljstvo korisnika. Ove vrste metrike usvajaju i kompanije i organizacije koje se odnose na standarde, kao što su Nacionalni institut za standarde i tehnologiju (u Sjedinjenim Državama) i Međunarodna organizacija za standardizaciju (ISO). Iako se ovi pokazatelji još uvijek često koriste i dobro prihvaćaju, oni su prikladni samo u situacijama kada se korištenje računara može podijeliti na specifične zadatke koji se sami mogu mjeriti kvantitativno i diskretno.

Shneiderman (2011) je opisao razliku između mikro i makro istraživanja interakcije čovjek - računar. Gore opisana istraživanja koja su fokusirana na korisničko iskustvo i koja podrazumijevaju dobro uspostavljene metrike i tehnike za poboljšanje performansi zadataka i vremena, mogla bi se smatrati istraživanjima mikro nivoa. Međutim, mnoge pojave koje u novije vrijeme zanimaju istraživače, kao što su motivacija, saradnja, društvena participacija, povjerenje i empatija, eventualni utjecaj na društvenom nivou, nije lako izmjeriti korištenjem metrika ili metoda sa mikro nivoa. Mnogi od ovih fenomena ne mogu se izmjeriti u laboratorijskim uslovima korištenjem istraživačkih modela iz psihologije tj. modela fokusiranih na zadatak. Istraživački nacrti koji su fokusirani na uspješnost rješavanja zadatka, ili vrijeme potrebno da se zadatak izvrši će nam malo dati podataka o recimo uživanju ili emocionalnoj dobiti. Postavlja se pitanje kako da mjerimo i objasnimo razloge zašto pojedinci koriste računare kada ne moraju ili zadovoljstvo poslom koji se obavlja isključivo uz pomoć računara. U ovim slučajevima potrebno je primijeniti više metoda za razumijevanje onoga što ove nove društveno-tehničke sisteme čini uspješnim. Jedan od ilustrativnih primjera fenomena korisničkog zadovoljstva koji se ne mogu jednostavno izučavati uz pomoć metrika uspješnosti i izvršenja zadatka je Wikipedia. Online enciklopedija ima više desetina hiljada aktivnih participanata koji odvajaju svoje vrijeme u kreiranju sadržaja, bez ikakve materijalne naknade.

Kao što je već spomenuto, istraživanja iz područja upotrebljivosti su nužno interdisciplinarna što sa sobom nosi značajne izazove. Dosta tih izazova proizilazi iz različitih istraživačkih potreba i očekivanja. Kada vodite interdisciplinarnu raspravu i radite na interdisciplinarnim timovima, važno je biti svjestan ovih razlika.

Ako posmatramo generalno, istraživanje upotrebljivosti zasnovano na sociologiji ima tendenciju da se fokusira na demografiju učesnika istraživanja i utvrđivanje da li su oni pravi slučajni uzorak. Navedeno se ne smatra relevantnim od strane IT stručnjaka, gdje se studenti informatike često koriste kao ispitanici (čak i kada to nije prikladno). Generalno, istraživanja upotrebljivosti temeljena na psihologiji imaju tendenciju da se fokusiraju na idealan i čist istraživački dizajn. Istraživanje zasnovano na kompjuterskoj nauci i dizajnu je više fokusirano na implikacije za tehničke osnove softvera ili na aspekt dizajna koji se više fokusira na izgled i iskustvo interfejsa (Lazar i sur. 2010).

Ovakva "podjela" istraživačkih pristupa u okviru istraživanja upotrebljivosti podrazumijeva generalizacije, koje opet nužno znače pojednostavljivanje radi ilustracije. Naravno da istraživači iz koje god discipline djelovali žele izvrsnost u svim fazama istraživanja. Ipak praksa pokazuje da se stvarno različite discipline intenzivnije fokusiraju na određene faze istraživanja. U prvom izdanju svoje knjige Lazar i sur. (2010) prikazuje fokuse različitih disciplina dijagramom koji se ne nalazi u drugom izdanju (Lazar i sur. 2017). Mi smo preuzeli i prilagodili taj dijagram (slika 1.1) svjesni zamki ovakve generalizacije, ali motivisani iskustvom rada u multidisciplinarnim istraživačkim timovima kada je razumijevanje motiva i fokusa stručnjaka drugih disciplina od izuzetne važnosti za uspješnu komunikaciju i zajednički rad.



Slika 1.1. Fokus različitih disciplina na određene aspekte procesa istraživanja

Ono što čini područje istraživanja upotrebljivosti sjajnim je upravo to što su sve istraživačke grupe koje sudjeluju ustvari neophodne i želimo da aspekt istraživačkog procesa koji izučavaju unaprijede maksimalno moguće. Za stvarni napredak područja istraživanja upotrebljivosti neophodne su različite prizme gledanja i različiti uglovi djelovanja. Rezultat koji imamo u praksi, a koji je direktna posljedica svega gore objašnjelog je u tome što istraživanja upotrebljivosti često podrazumijevaju korištenje više istraživačkih postupaka.

Sauro (2016) navodi da ne postoje konkretne smjernice o tome kako izmjeriti upotrebljivost tj. njene aspkete kao što su efektivnost, efikasnost, korisničko zadovoljstvo ili neke druge varijable makro nivoa, upravo zbog kompleksnosti interdisciplinarnog pristupa koji je neophodan. Ipak meta analize daju odgovor na pitanje šta istraživači najčešće koriste. Većina istraživanja sadrži neku kombinaciju mjera kao što je: stope uspješnog završetka zadatka, broja napravljenih grešaka, vremena potrebnog za izvođenje zadataka, zadovoljstva prilikom obavljanja zadatka, pristupa korisničkoj podršci i popisa problema s upotrebom. Ovakvi nalazi su u skladu sa prethodno opisanim raznolikim pitanjima koja se postavljaju u savremenim istaživanjima korisničkog iskustva.

Kada je tema istraživanja nova potrebno je razmišljati i započeti s istraživačkom metodom koja nam daje puno (uglavnom nestruktuiranih) informacija – kao što su ankete sa otvorenim pitanjima, intervjui, fokus grupe i etnografija. Zatim, na osnovu razumijevanja iz nekoliko istraživačkih studija, mogu se izvesti istraživačke studije koje koriste strukturiranije istraživačke metode—kao što su eksperimentalni dizajn, automatizirano prikupljanje podataka i vremenski dnevnic.

Općenito postoje dvije vrste testova upotrebljivosti: pronalaženje i rješavanje problema upotrebljivosti (formativni testovi) i opis upotrebljivosti aplikacije pomoću metrike (sumativni testovi). Izrazi formativno i sumativno potječu iz obrazovanja gdje se na sličan način koriste za opisivanje testova učenja učenika: formativni - pružajući neposredne povratne informacije za poboljšanje učenja, nasuprot sumativnim - ocjenjivanjem naučenog.

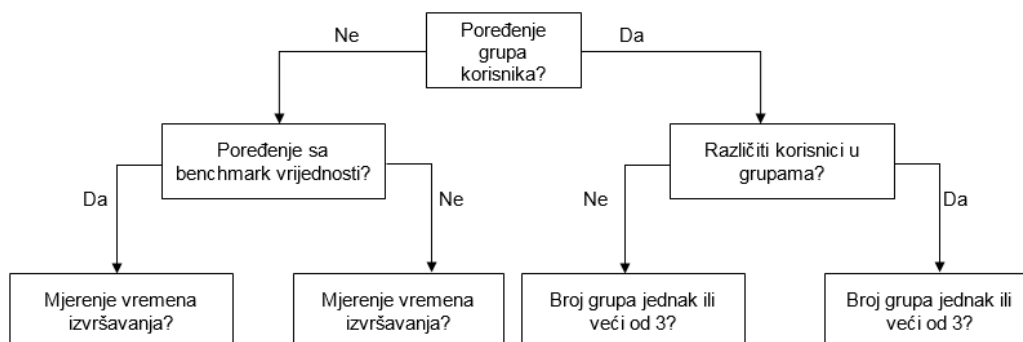
Kod formativnih istraživanja uglavnom se koriste mali uzorci, a podaci se analiziraju kvalitativno. Podaci koji se analiziraju uglavnom sadrže opis problema i preporuke za unapređenje dizajna. Argument za ovakav pristup je u tome da je sam cilj testiranja upotrebljivosti ustvari pronaći i riješiti što više problema, čime se aplikacija unaprijeđuje. Ipak to ne znači da nema mogućnosti kvantifikacije rezultata. Moguće je kvantificirati probleme u smislu učestalosti i ozbiljnosti (analiza frekvencija), pratiti koji su se korisnici

susreli s kojim problemima, izmjeriti koliko im je vremena trebalo da izvrše zadatke i utvrditi jesu li ih uspješno završili.

Postoje dvije vrste sumativnih testova: referentni i uporedni. Cilj mjerenja upotrebljivosti na ovakav način je opisati koliko je aplikacija upotrebljiva u odnosu na skup referentnih ciljeva. Za ove referentne ciljeve, odnosno vrijednosti, koristi se engleski naziv „benchmark“, a testovi kada se izmjerene metrike porede sa benchmark vrijednosti zovu se benchmark testovi. Benchmark testovi pružaju ulazne podatke o tome šta ispraviti u okviru same aplikacija, a takođe pružaju osnovnu osnovu za poređenje nakon načinjenih izmjena.

Komparativni test upotrebljivosti, kao što i samo ime govori, uključuje više od jedne aplikacije. To može biti usporedba nove s prethodnom verzijom proizvoda ili usporedba konkurentskih proizvoda. Kod komparativnih testova, isti korisnici mogu pokušati izvršiti zadatke na svim proizvodima (dizajn unutar predmeta) ili različiti skupovi korisnika mogu raditi sa svakim proizvodom (dizajn između predmeta).

Svrha ove knjige je da pruži praktične upute i rješenja prilikom dizajniranja istraživanja korisničkog iskustva. U nastavku na slici 1.2. nudimo primjer algoritma donošenja odluke o tome koji tip studije i koji nacrt koristiti i koji statistički postupci logički i znanstveno prate takvo istraživanje (Sauro 2016).



Slika 1.2. Algoritam odlučivanja o tipu studije (prema Sauro 2016)

Ovaj primjer ilustruje algoritam odlučivanja, koji ovdje nije potpun, i u sljedećim poglavljima ćemo uvesti vezu određenih eksperimentalnih nacrtu sa odgovarajućim statističkim metodama, kao i druge algoritme odlučivanja.

Kako bismo izbjegli stvaranje konfuzije kod čitaoca ovdje ćemo objasniti termin varijabla koji će se u nastavku teksta koristiti. Pod varijablom podrazumijevamo osobinu ispitanika ili situacije koja može poprimiti različite vrijednosti. Suprotan pojam od varijable je konstanta. Konstanta predstavlja karakteristiku koja može imati samo jednu vrijednost (npr. broj glava kod psa).

U nastavku, pored objašnjenja osnovnih karakteristika i razlika između istraživačkih nacrtu, te primjera istraživanja, ponudit ćemo osnovne informacije o alatima za prikupljanje podataka koji se najčešće koriste: Likertova skala i semantički diferencijal. Također, sva objašnjenja će pratiti primjeri, kako iz područja društvenih znanosti, tako i iz područja istraživanja upotrebljivosti.

2. Odabir ispitanika u istraživanju

Za početak, potrebno je definirati osnovne parametre relevantne za odabir ispitanika u istraživanju. Pitanje od kojeg najčešće počinjemo je: „Ko su pojedinci na koje želimo da generalizujemo svoje nalaze?“ Razlog za neophodnost ovakvog pristupa leži u tome što smo uvijek zainteresirani za širu skupinu pojedinaca od one na kojoj možemo realno i praktično realizirati istraživanje. U istraživanjima upotrebljivosti uglavnom imamo ciljnu skupinu korisnika proizvoda npr. studente kojima bi naš program bio vrlo koristan alat u praktičnim vježbama ili uposlenici javne uprave koji bi trebali da prihvate novi informacijski sistem. Ispitivanje upotrebljivosti na svim jedinkama je skupo, nepraktično i u konačnici za dobijanje odgovora koji su nama potrebni vezano za proizvod i nepotrebno. Ipak, potrebno je ispuniti neke minimalne kriterije kada je odabir ispitanika te njihov broj u pitanju kako bi mogli donositi valjane zaključke.

Sam način odabira ispitanika u uzorak je proces niza logičnih pitanja i odgovora koje dajemo vezano za populaciju koja nas zanima i konkretne operativne ciljeve istraživanja. Odluka o broju ispitanika koji će sudjelovati u istraživanju zavisi opet od ciljeva tj. pitanja na koja želimo dobiti odgovor svojim istraživanjem i sa tim usko vezano je pitanje statističkih postupaka koji se trebaju koristiti kako bi se dobili željeni odgovori. Prije nego uđemo u daljnju elaboraciju odabira ispitanika i odluke o broju potrebno je definirati osnovne pojmove.

Populacija se definira kao skup stvari, događaja ili bilo kojih entiteta na koje želimo da se odnose naše generalizacije. Pojedince nazivamo članovima populacije. U situaciji kada želimo svoje rezultate generalizirati na recimo pripadnike određene dobne skupine, ili sve osobe koje se bave određenim zanimanjem istraživanje na populaciji je nemoguće i nepraktično.

Fajgelj (2005) navodi da je uzorak „staro i dobro zdravorazumsko“ rješenje problema za nemogućnost izvođenja istraživanja na populaciji od interesa, ustvari to da se istraživanje provede na uzorku članova populacije. Prilikom uzorkovanja mi preuzimamo određeni rizik da ono što smo saznali na uzorku nužno i u totalu možemo primijeniti na populaciju. Ipak, potrebno je ispoštovati metodološki definirane korake i pretpostavke kako bismo mogli ocijeniti vjerovatnoću greške prilikom generalizacije rezultata.

Reprezentativnost uzorka je jedan od osnovnih parametara kvaliteta uzorka. Za reprezentativan uzorak kažemo da je „populacija u malom“ s obzirom na sve relevantne karakteristike populacije. Ovdje je potrebno objasniti na šta mislimo kada kažemo relevantne karakteristike populacije. Uzmimo za primjer istraživanje upotrebljivosti novog računovodstvenog programa. Populacija od interesa za ovo istraživanje su sve jedinice koje bi taj softver trebale koristiti. Prilikom planiranja istraživanja važno je razmisliti koje su karakteristike jedinice u ovoj populaciji koje su nam relevantne za istraživanje. Npr. bitna nam je dob i godine radnog iskustva jer je moguće da početnici i iskusni korisnici različito doživljavaju i upotrebljavaju proizvod. Spolna struktura je jedna od varijabli koje uvijek

gledamo uzeti u obzir zbog postojanja što stečenih što naučenih spolnih razlika u znanjima i vještinama. Bitan nam je i nivo i vrsta obrazovanja jedinki u populaciji s obzirom na to da vjerovatno utječu na radni učinak. S druge strane, broj članova domaćinstva nije varijabla koja je relevantna za ovo istraživanje jer nema logičnog uporišta za pretpostavku da bi mogla utjecati na rezultate. Dakle, kada definiramo uzorak, važno je da je on reprezentativan za populaciju od interesa s obzirom na relevantne karakteristike populacije. Ova analiza relevantnih karakteristika populacije je jedan od ključnih elemenata planiranja istraživanja. Problem se javlja kada se dobiveni rezultati mogu objasniti ne promjenama uzrokovanim našim predmetom mjerenja nego nekom karakteristikom ispitanika u našem uzorku koju nismo uzeli u obzir.

U istraživanjima upotrebljivosti nije uvijek neophodno imati reprezentativan uzorak. Odluka o tome kakav uzorak je potreban zavisi i od faze razvoja proizvoda u kojoj se proizvod testira. U fazi razvoja proizvoda relevantna nam je informacija od korisnika općenito jer informacije koje dobijemo koristimo kao svojevrsan putokaz za daljnji razvoj. U finalnoj fazi pripreme proizvoda potrebne su nam rafiniranije informacije od većeg broja ispitanika. Iz tog razloga u ovoj fazi je neophodno povesti više računa o reprezentativnosti uzorka.

Pitanje veličine uzorka se mora adresirati i u kontekstu vrste istraživanja upotrebljivosti koju primjenjujemo (Sauro 2016). Ukoliko koristimo sumativni pristup, onda nas zanima samo procenat uspješnosti završetka zadatka, a ukoliko koristimo formativni pristup onda ustvari radimo dijagnostičku evaluaciju u cilju unapređenja različitih aspekata softvera.

Veličina uzorka je neophodan ali nedovoljan uslov za postizanje reprezentativnosti. Dakle, uzorak koji je mali nije reprezentativan, a veliki uzorak može i ne mora biti reprezentativan. Kako je ovo moguće? Uzmimo za primjer da razvijamo studentski online servis za studente Univerziteta u Sarajevu. Populacija od interesa su nam svi studenti koji studiraju na Univerzitetu u Sarajevu. Uzorak u studiji upotrebljivosti koji bi činilo 1000 studenata Filozofskog fakulteta je po svim statističkim parametrima velik, ali za ovu populaciju nije reprezentativan. Mjeriti upotrebljivost samo na studentima Filozofskog fakulteta ima ozbiljne nedostatke, a svi proizilaze iz činjenice da su studenti Filozofskog fakulteta podskupina populacije sa za sebe karakterističnim osobinama. Naravno da reprezentativan uzorak studenata Univerziteta u Sarajevu mora sadržavati i studente Filozofskog fakulteta, ali ne smije sadržavati samo te studente. Razlog za ovo leži u činjenici da se studenti Filozofskog fakulteta vjerovatno razlikuju u načinu i vještini prilikom korištenja IT opreme u odnosu na npr. studente Elektrotehničkog fakulteta ili Mašinskog fakulteta, pa i studenata Akademije likovne umjetnosti ili Muzičke akademije.

Generalno, kada nas istraživanje ne bi ništa koštalo ili da nemamo ograničeno vrijeme provođenja naših studija, uvijek bismo imali velike uzorke. Sa povećanjem veličine uzorka mogućnost greške u zaključivanju se smanjuje, što posljedično znači da smo sigurniji u rezultate koje smo dobili. Primjer ovakvih istraživanja (koja ne koštaju značajno više a može se imati veliki uzorak) su web bazirane ankete koje se distribuiraju putem e-maila ili društvenih mreža. Angažovanje većeg broja demografski različitih ispitanika je moguće i korištenjem online web servisa namijenjenih za istraživanja. Primjer je Amazonova platforma Mechanical Turk (Buhrmester i sur. 2011) koja se sve više koristi za evaluacije jer omogućava

jednostavan način angažovanja velikog broja različitih korisnika. Naravno, opredjeljenje da se koristi web servis ovog tipa podrazumijeva da je interfejs aplikacije koja se evaluira na engleskom jeziku, kao i upitnici koji se koriste u istraživanjima.

Naravno, potrebno je imati i neke kvantitativne odrednice veličine uzorka. Fajgelj (2005) navodi da se statistički svaki uzorak koji je manji od 30 tretira kao mali, uzorci između 30 i 100 srednjim, a uzorci sa više od 100 ispitanika se smatraju velikim. Odluka o neophodnoj veličini uzorka da bi se donosili relevantni zaključci ovisi o cilju mjerenja (kolika nam je preciznost neophodna prilikom zaključivanja), broju varijabli i vrsti analize koja se radi. Generalno pravilo je da je optimalan omjer varijabli i ispitanika 1:5, a minimalan 1:3. Ipak, prilikom analize različitih nacrtâ istraživanja spomenut ćemo konkretnije parametre za određivanje veličine uzorka.

Kod evaluacije korisničkog iskustva za određenu aplikaciji ciljna populacija korisnika može biti poznata, a primjer je kada se implementira softver za knjigovodstvo za određenu firmu. Kod usluga online plaćanja za vlasnike računa neke banke, populacija je značajno veća, ali je također definisana i banka – kao vlasnik softvera - raspolaže podacima o demografskim osobinama korisnika poput spola, godina, obrazovanja. Populacija korisnika za aplikacije poput MS Office 365 je manje definirana i više se pažnje treba posvetiti reprezentativnosti prilikom odabira ispitanika. U slučaju kada ciljna populacija nije dobro definirana, onda se treba predvidjeti uključanje većeg broja ispitanika i to tako da se uključe predstavnici prepoznatih podgrupa unutar populacije.

Jedan od načina da se utvrdi validnost odgovora ispitanika je da se od ispitanika prikupe demografski podaci da se osigura da odgovori predstavljaju raznolik, poprečni presjek ispitanika ili da su odgovori dovoljno reprezentativni za već utvrđene osnovne podatke o populaciji: dob, spol, obrazovanje, odgovornosti na poslu, predznanje korištenja računara (Lazar i sur. 2017). Ovo je posebno važno kada se ispitanici biraju pozivom, a ne slučajnim odabirom, a prezentacija prikupljenih demografskih podataka obavezno prati prezentaciju rezultata istraživanja.

Ovdje je potrebno napomenuti da je reprezentativnost uzorka ključan uslov za postizanje vanjske valjanosti istraživanja. Vanjska, ili ekološka valjanost istraživanja se odnosi na mogućnost generaliziranja rezultata koje smo dobili u svom istraživanju na pripadnike populacije od interesa i na kontekst tj. situacije koje su za naš problem relevantne. Generalno, što su više artificijelni uslovi u kojima provodimo istraživanje, tj. što više varijabli kontrolišemo to je manja vanjska valjanost. Također, ukoliko koristimo nereprezentativan uzorak nećemo moći svoje rezultate generalizirati na kompletnu populaciju, što u suštini znači nisku vanjsku valjanost.

Iako se evaluacija upotrebljivosti značajno razlikuje od drugih istraživanja, etička pravila vezana za rad sa ljudskim subjektima su ista: učesnici moraju biti obaviješteni o svojim obavezama i pravima, o ciljevima istraživanja i načinu korištenja prikupljenih podataka, moraju odobriti snimanje odogovora i video zapisa, trebaju imati dozvolu da napuste istraživanje bilo kada; i da, osim ako izričito nisu dali dozvolu, njihovo učešće ostaje anonimno i njihovi podaci ostaju zaštićeni.

U slučaju evaluacije upotrebljivosti aplikacija virtualne realnosti moguće su neželjene pojave poput dezorijentacije, vrtoglavice i mučnine, tako da je tada potrebna posebna obazrivost evaluatora i posmatranje korisnika tokom cijelog trajanja evaluacije.

3. Mjerenje i mjerni instrumenti

U kontekstu mjerenja upotrebljivosti mjerni instrument je alat za prikupljanje informacija od ispitanika. Postoje različite vrste mjernih instrumenata koje je moguće implementirati u istraživanjima. Prije elaboracije o vrstama mjernih instrumenata, potrebno je objasniti šta je mjerenje kao i šta su i zašto su važne mjerne karakteristike bilo kog instrumenta. Nakon toga ćemo se fokusirati na upitnike i standardizirane skale procjene kao najčešće korištene u ispitivanjima efikasnosti.

3.1. Mjerenje

Kako bismo mogli u nastavku na adekvatan način govoriti o mjerenju upotrebljivosti, neophodno je da u prvom redu objasnimo koncept mjerenja u društvenim/humanističkim naukama, kao i neke od najčešćih problema i zahtjeva svakog mjerenja pa i mjerenja upotrebljivosti. Još je Stevens (1946) objasnio da skala mjerenja zavisi od vrste empirijskih operacija koje se mogu izvoditi s podacima. Ove operacije limitirane su prirodom karakteristike koja se mjeri / skalira i odabirom procedura. U nastavku ćemo pobliže objasniti četiri skale mjerenja: nominalnu, ordinalnu, intervalnu i omjernu.

a) Nominalna skala mjerenja

Nominalna skala predstavlja najmanje restriktivan način dodjeljivanja brojeva. U slučaju nominalne skale, brojevi se dodjeljuju kao simboli ili opisi i mogu biti zamijenjeni bilo kojim drugim simbolima, recimo slovima ili riječima. Na primjer, u jednoj školi odjeljenja su numerirana brojevima: 1,2,... (primjer: VIII₁, VIII₂), a u drugoj slovima: a,b,... (primjer: VIII_a, VIII_b). Suština korištenja broja je samo da se označi grupa. S obzirom na to da je svrha obilježavanja zadovoljena i u situaciji kada obilježja zamijene mjesta (na primjer, VIII₁ postane VIII₂, a VIII₂ postane VIII₁) ova skala ostaje invarijantna prilikom zamjene ili permutacije brojeva kao simbola. Nominalna skala je najjednostavnija forma mjerenja, a neki autori smatraju da nije primjereno ovo zvati mjerenjem. Međutim, ako je mjerenje "dodjeljivanje brojeva prema nekom pravilu", onda poštivanje pravila "ne smije se dodijeliti isti broj kao oznaka za dvije različite grupe" kvalificira nominalnu skalu kao mjerenje.

b) Ordinalna skala mjerenja

Ordinalna skala proizlazi iz operacije „redanje po veličini“. Kako nijedna transformacija koja čuva rangove neće utjecati na variranje, ova skala se može zvati i izotonična. Ordinalna skala daje informaciju o poretku objekata u nekoj grupi, ali ne postoji jednakost u razlikama među rangovima. Primjer je mjesto osvojeno u trci, na takmičenju. Prvo, drugo, treće i četvrto mjesto nam daju informaciju o redoslijedu takmičara, ali nemamo informaciju kolike su razlike među njima u vremenu ili bodovima. Značajna ordinalna skala

koja se koristi u mjerenjima korisničkog iskustva i upotrebljivosti je Likertova skala slaganja, o čemu će biti više informacija dato u nastavku knjige.

c) Intervalna skala mjerenja

Intervalna skala nam daje informaciju o poretku, posjeduje jednakost razlika među intervalima, ali nema apsolutne nule. Nula na intervalnoj skali je arbitrarno određena, odnosno rezultat 0 ne znači da neki atribut uopće ne postoji, nego da njegovo prisustvo nije zahvaćeno tom metodom mjerenja. Primjeri intervalne skale u mjerenju su skale za mjerenje temperature u stepenima Celzijuma i stepenima Farenhajta. Na primjer, temperatura od 30°C je 88 stepeni Farenhajta. Da je ovo omjerna skala, onda bi temperatura od 15 stepeni Celzija (upola manja od početne) bila 44 stepena Farenhajta, a to nije tako.

d) Omjerna skala

Omjerna skala, koju uglavnom nalazimo u mjerenjima u prirodnim znanostima, je skala koja ima zadovoljene sve uvjete: veličinski odnos, jednakost intervala i jednakost omjera, te apsolutnu nulu. Ovo je skala s kojom se mogu raditi sve statističke operacije, rezultati se mogu transformirati tako da ih samo pomnožimo s nekom konstantom. Nula na omjernoj skali znači da neko svojstvo ne postoji (čak i onda kada se to nikada nije desilo, ali se može desiti). Omjerna skala su visina i težina ispitanika, vrijeme za koje su trkači istrčali utrku, itd.

Sumarni pregled skala, njima odgovarajućih osnovnih empirijskih operacija i primjenjivih statistika je dat u tabeli 3.1. Treba istaknuti kumulativnu prirodu kolone u kojoj se navode osnovne operacije potrebne za kreiranje svake vrste skale i kolone sa primjenjivim statistikama. To znači da na operaciju navedenu nasuprot određenoj skali moraju se dodati sve operacije navedene za skale koje joj prethode. Isto vrijedi i za nevedene statistike, i kako je kolona kumulativna sve statistike navedene u tabeli su primjenjive za podatke omjerne skale.

Tabela 3.1. Pregled skala, empirijskih operacija i primjenjivih statistika po Stivensu (1946)

Skala	Osnovne empirijske operacije	Primjenjive statistike
Nominalna	Određivanje jednakosti	Broj, proporcija, mod
Ordinalna	Određivanje odnosa veće ili manje	Medijan, percentili
Intervalna	Određivanje jednakosti ili razlike intervala	Srednja vrijednost, standardna devijacija
Omjerna	Određivanje jednakosti omjera	Koeficijent varijacije

3.2. Greške u mjerenju

Pogreška mjerenja je razlika između stvarnog rezultata ispitanika na nekom testu i onoga koji smo mi zabilježili. Pogreške mjerenja dijelimo u dvije osnovne grupe:

1. Slučajne pogreške ili pogreške koje nastaju uslijed djelovanja nesistematskih faktora. Postoje dvije ključne karakteristike kako nesistematski faktori djeluju na rezultat: a) djeluju po slučaju tako da se vjerovatnoća pogreške mijenja po slučaju; b) djeluju u oba smjera tako da nekada povećavaju, a nekada smanjuju rezultat.

2. Sistematske pogreške ili pogreške koje nastaju zbog djelovanja sistematskih faktora. Sistematske pogreške u mjerenju javljaju se pod utjecajem varijabli koje na rezultate djeluju tako da ih mijenjaju kod svih ispitanika u istom smjeru. Ove varijable mogu biti namjerno manipulirane/ uvedene kako bi se odredio utjecaj te varijable na rezultate ispitanika. Sistematski faktori su predvidljiviji i njima ili manipuliramo (u slučaju eksperimentalnih varijabli) ili ih nastojimo držati pod kontrolom za vrijeme ispitivanja.

3.3. Karakteristike mjernih instrumenata

Bilo da sprovodimo eksperimentalno istraživanje, korelacijsko ili anketno istraživanje, pitanje mjernog instrumenta je jedno od centralnih. Naime, od kvalitete primjenjenog instrumenta ovisi mogućnost donošenja relevantnih zaključaka u istraživanju. Prilikom analize metrijskih karakteristika instrumenta razmatramo valjanost, pouzdanost, objektivnost i diskriminativnost.

3.3.1. Valjanost

Najjednostavniji način za objasniti valjanost je na način da postavimo sebi pitanje: „Da li ovaj instrument stvarno mjeri ono što mislimo da mjeri?“. Da stvar bude kompliciranija, moguće je da je jedan instrument valjan u jednoj, a potpuno neprikladan u drugoj situaciji. S tim u vezi, procjenu valjanosti obično dijelimo na procjenu sadržajne, kriterijske i konstruktne valjanosti.

Milas (2005) navodi kako je sadržajna valjanost stepen u kojem upitnik svojim sadržajem, zadacima, pitanjima i drugim jedinicama tematski prilagođen temi koju ispituje. Npr. kada govorimo o testovima znanja iz nekog predmeta koji trebaju da evaluiraju učeničko znanje nakon polugodišta, sadržajno valjan test bi sadržavao pitanja iz svih nastavnih jedinica, a ne samo nekih koje nastavnik cijeni najvažnijim. Slično, prilikom evaluacije upotrebljivosti, sadržajno valjan upitnik bi se fokusirao na sve aspekte upotrebljivosti kao i na sve dijelove proizvoda koji se treba evaluirati. Kako bismo provjerili ovaj aspekt valjanosti najčešće koristimo procjenu stručnjaka iz područja koje istražujemo, npr. stručnjaka za razvoj edukacijskog softvera.

Kriterijska valjanost podrazumijeva povezanost rezultata na našem upitniku sa nekim vanjskim kriterijem. Vanjski kriterij je varijabla sa kojom očekujemo na osnovu logike, teorijskih razmatranja ili rezultata prethodnih istraživanja povezanost varijable koju mjerimo. Kriterijska valjanost se brojčano izražava koeficijentom korelacije tj. povezanosti varijable koju mjerimo i kriterijske varijable. Što je koeficijent korelacije bliži ± 1 to je korelacija veća i time bolja naša kriterijska valjanost. Kod mjerenja efikasnosti treninga trebalo bi mjeriti volumen potrošnje kisika VO₂ što je zahtjevna procedura, ali se kao kriterijska varijabla koristi srčani ritam koji se jednostavno mjeri korištenjem različitih nosivih elektronskih

uređaja tokom treninga. Napomena: povezanost varijabli iz ovog primjera je različita kod različitih nivoa opterećenja i to se treba uzeti u obzir.

Razlikujemo dijagnostičku i prognostičku kriterijsku valjanost. Kod evaluacije dijagnostičke kriterijske valjanosti obje varijable su mjerene u istoj vremenskoj tački ili vremenski blizu. Evaluacija prognostičke valjanosti podrazumijeva analizu povezanosti između vremenski odvojenih varijabli. Kada govorimo o valjanosti upitnika upotrebljivosti softvera jedan od načina je recimo mjeriti korelaciju između rezultata na upitniku upotrebljivosti i vremena provedenog u korištenju programa/aplikacije, ili broja aktivnosti koje su urađene koristeći usluge koje softver nudi. Također, kada govorimo o obrazovnom softveru relevantna mjera valjanosti je i ishod učenja npr. rezultat na testu znanja.

Konstruktna valjanost je vrsta valjanosti koja se ustvari i reflektira kroz već spomenuti upit o tome da li instrument zaista mjeri ono što mi mislimo da mjeri. Jedan od načina utvrđivanja konstruktne valjanosti je ispitivanje korelacije između našeg i instrumenta koji mjeri isti fenomen. Naravno da očekujemo izuzetno visoku korelaciju između tih rezultata. Optimalan način primjene ove tehnike podrazumijeva ispitivanje korelacija sa četiri mjerenja, gdje će dva mjerenja biti mjerenje istog konstrukta a druga dva ista metoda. Pored ovog načina konstruktnu valjanost možemo provjeriti i postupkom faktorske analize koja nam otkriva zasićenost rezultata na upitniku latentnim varijablama.

3.3.2. Pouzdanost

Milas (2005) definira pouzdanost kao odsutnost greške mjerenja i dosljednost mjernog postupka, tj. stupanj ponovljivosti rezultata. Procjena pouzdanosti se vrši na dva načina: testiranju vremenske konzistencije tj. ponovljenim mjerenjima i testiranju unutarnje konzistencije instrumenta. Kada govorimo o testiranju sa ponovljenim mjerenjima, obično primjenjujemo paralelne forme instrumenta i uz pomoć koeficijenta korelacije između tih mjerenja donosimo zaključke. Pouzdanost tipa unutarnje konzistencije se najčešće izražava **Cronbach α** koeficijentom pouzdanosti. Ovaj koeficijent može imati vrijednosti od 0 do 1, gdje je 1 idealna, teorijska pouzdanost.

3.3.3. Diskriminativnost

Diskriminativnost se odnosi na osobinu instrumenta da razlikuje ispitanike s obzirom na predmet mjerenja. Plastičan primjer za diskriminativnosti je pitanje mjerenja znanja testom znanja. Test znanja iz nekog područja koji „svi zadovolje“ ili „niko ne zadovolji“ je potpuno nediskriminativan. U tom smislu ne služi svojoj svrsi razlikovanja ispitanika s obzirom na znanje iz određenog područja. Nemoguće je da svi ispitanici imaju isti nivo znanja, što znači da instrument nije dovoljno osjetljiv da detektira razlike. Sličan problem se javlja i u testiranju upotrebljivosti kada dobijemo nalaz da su „svi zadovoljni“ ili „sve je super“. U slučaju upotrebljivosti ovakvi rezultati mjerenja se okarakterišu kao mjere taštine (eng. *vanity metrics*) jer su korisni samo za taštinu kreatora proizvoda.

3.3.4. Objektivnost

Objektivnost mjerenja općenito podrazumijeva pojavu da rezultati mjerenja ne zavise od samog mjernog postupka i instrumenata. Drugim riječima, da ukoliko želimo da na drugi

način sa drugim istraživačima izmjerimo neku varijablu, možemo to učiniti i dobiti iste ili približne rezultate.

3.4. Reprezentativnost uzorka kod mjerenja

Donošenje relevantnih zaključaka o populaciji u istraživanju zavisi od kvalitete primjenjenog instrumenta, ali i od uzorka nad kojim se instrument mjerenja primjenjuje i koji je izabran da predstavlja populacije koji su uključene u proces mjerenja. U prethodnom poglavlju u kojem smo se dotakli pitanja odabira uzorka naglasili smo važnost reprezentativnosti uzorka, kao i to da često u istraživanjima upotrebljivosti, u zavisnosti od njihovog cilja, od reprezentativnosti odustanemo. Ukoliko pravimo anketno istraživanje kako bismo prikupili obimnije i iscrpnije podatke vezane za neki softverski proizvod koji nas zanima, onda nam je (kao i svim istraživačima koji prikupljaju podatke putem anketiranja) reprezentativnost važna. Iz tog razloga će konkretni postupci uzorkovanja sa ciljem postizanja reprezentativnosti biti ovdje objašnjeni.

Milas (2005) navodi niz koraka, ili kako on naziva „praktičnih problema koje treba riješiti“, prilikom uzorkovanja: utvrđivanje ciljne populacije, utvrđivanje okvira uzorkovanja, izbor metode uzorkovanja i određivanje potrebne veličine uzorka. U nastavku su izložene i diskutovane moguće odluke prilikom ovih izbora.

3.4.1. Utvrđivanje ciljne populacije

Jasna definicija populacije koja nas zanima je vrlo važan jedan od početnih koraka u istraživanju. Definiranjem ciljne populacije trebamo tačno ograničiti koji pojedinci dolaze u obzir za naše istraživanje, a koji ne. Naizgled radi se o jednostavnom zadatku, ali ukoliko uđemo dublje u suštinu vidjet ćemo da je potrebno uzeti u obzir više faktora nego što se na prvi pogled čini. Npr. ukoliko nas zanimaju svi punoljetni građani Kantona Sarajevo, potrebno je da definiramo da li su to npr. građani sa prijavljenim prebivalištem u Sarajevu (a koji mogu živjeti i u inozemstvu, trajno ili privremeno), ili su to osobe koje borave u Sarajevu već određeni period (definirati koji je to vremenski period), potrebno je naznačiti od kojeg datuma minimum osoba treba biti punoljetna (šta ćemo sa onima koji postanu punoljetni u toku istraživanja i sl). Dakle definiranje ciljne populacije je korak u kojem moramo biti vrlo precizni kako ne bismo morali donositi ad-hoc odluke u procesu istraživanja.

3.4.2. Utvrđivanje okvira uzorkovanja

Okvir uzorkovanja u biti podrazumijeva najbolju dostupnu operacionalizaciju populacije. U idealnom slučaju govorimo o spisku pripadnika populacije od interesa, iako je to rijetko slučaj. Najčešći je problem sa dostupnosti podataka, kao i sa cinjenicom da se populacija od interesa u većini slučajeva stalno mijenja (ljudi umiru, mijenjaju mjesto prebivališta i sl.). Vrlo je važno definirati okvir uzorkovanja, jer ukoliko odaberemo pogrešan okvir, može doći do značajnog iskrivljivanja uzorka. Najpoznatiji je primjer časopisa Literary digest sa kraja 19. stoljeća koji je pokušao anketom predvidjeti ishod predsjedničkih izbora u SAD-u. Oni su poslali upit ogromnom broju ispitanika vjerujući kako će time osigurati reprezentativnost uzorka i tačnije podatke. Spisak populacije od interesa (tj. okvir uzorkovanja) su uzeli iz tada dostupnog telefonskog imenika i liste vlasnika registrovanih

automobila. Rezultati anketiranja su govorili u prilog uvjerljive pobjede republikanskog kandidata. Ishod izbora je bio potpuno suprotan i ustvari je demokratski kandidat odnio jednu od najuvjerljivijih pobjeda u historiji američkih izbora. Uzrok ovako velike greške u zaključivanju upravo možemo tražiti u definiranom okviru uzorkovanja. Naime, u tom periodu, stanovništvo koje je posjedovalo telefonski aparat u kući i automobil je pripadalo bogatijoj populaciji, koja je u SAD-u dominantno republikanski orijentisanja. Ovakvim odabirom okvira uzorkovanja u iskrivili smo uzorak koji više nije mogao biti reprezentativan bez obzira na veličinu.

3.4.3. Izbor metode uzorkovanja

Postoje dvije kategorije načina na koje možemo provesti postupak uzorkovanja: probabilistički i neprobabilistički. Osnovna razlika između ovih pristupa se ogleda u tome da li svaki član ima poznatu, nenultu vjerovatnoću da uđe u uzorak (probabilistički način) ili je ta vjerovatnoća nula/nepoznata (neprobabilistički način uzorkovanja).

1) Neprobabilistički način istraživači češće biraju jer je 1) lakše doći do uzorka 2) lakše je formirati uzorak na način da odgovori na specifične potrebe istraživanja. Iz ove dvije prednosti proizilazi i veća ekonomičnost tj. manji troškovi primjene ovog načina uzorkovanja, jednostavnost primjene i kraći period sprovođenja istraživanja. Nedostatak neprobabilističkih uzoraka je u činjenici da s obzirom na to da nije poznata vjerovatnoća svakog pripadnika populacije da uđe u uzorak, ne možemo tvrditi da je uzorak koji primjenjujemo reprezentativan tj. da je nepristrasan.

Najčešće korištena vrsta uzorka, i neprobabilističkog uzorka je prigodni uzorak. Prigodni uzorak je dio populacije od interesa koji nam je dostupniji od ostalih. Npr. to su studenti našeg fakulteta u odnosu na studente drugih fakulteta/ univerziteta. Prigodni uzorak se toliko često koristi, i to baš u ovom obliku, da su neki kritičari moderne znanosti, znanost današnjeg doba okarakterizirali kao znanost o ponašanju studenata (recimo studenata psihologije). Ovdje ipak potrebno napomenuti da su recimo eksperimentalna i anketna istraživanja različite vrste istraživanja sa različitim istraživačkim ciljevima. Sa tim u vezi je i važnost reprezentativnosti tj. načina izbora ispitanika različita. U eksperimentalnim istraživanjima insistiramo na unutarnjoj valjanosti, gdje nam je ključno pitanje izjednačenosti grupa i uslova, dok je u anketnim istraživanjima reprezentativnost bitna kako bismo imali zadovoljavajuću vanjsku valjanost tj. mogućnost generalizacije rezultata na cijelu populaciju.

Pored prigodnosti uzorka, često u istraživanjima koristimo dobrovoljce, dakle pojedince koji se sami javljaju da sudjeluju u istraživanjima. Analize i istraživanja dobrovoljaca su pokazala da se dobrovoljci od ostatka populacije značajno razlikuju te je mogućnost uopćavanja rezultata koje smo dobili na dobrovoljcima vrlo upitna. Milas (2005) navodi kako se dobrovoljci od ne-dobrovoljaca razlikuju po obrazovanju, društvenoj klasi, inteligenciji, potrebi za odobravanjem i društvenosti. Dakle, u prosjeku, dobrovoljci su osobe višeg obrazovanja, pripadnici više društvene klase, inteligentniji, društveniji i imaju veću potrebu za odobravanjem. Jedna od hipoteza koja se može postaviti na osnovu ovih informacija je ta da dobrovoljci lakše mogu da pretpostave istraživačke hipoteze te da su zbog svoje potrebe za odobravanjem skloniji da svoje ponašanje prilagode percipiranim očekivanjima.

Kvotni uzorak je vrsta neprobabilističkog uzorka u kojem se unaprijed definiraju podgrupe populacije na osnovu varijabli značajnih za mjerenje. Primjer za kvotni uzorak je recimo definiranje kvota unutar populacije s obzirom na nivo obrazovanja. Prva teškoća koja se javlja prilikom primjene kvotnog uzorka je vezana za definiranje kvota. Mi kako bismo definirali kvote moramo imati tačnu informaciju o veličini ciljne populacije, kao i o brojnosti tj. udjelu svake od podskupina na koje smo populaciju podijelili. Očiti praktični problem koji se javlja je vezan za definiranje brojnosti populacije i podskupina unutar populacije s obzirom na činjenicu da se ti brojevi stalno mijenjaju na svim navedenim nivoima. Npr. ukoliko uzmemo rezultate popisa stanovništva BiH koji je provđen 2013. izlažemo se mogućnosti velike greške s obzirom na protok vremena od posljednjeg popisa u kojima su se vjerovatno desile značajne promjene smislu broja stanovnika sa određenim nivoom obrazovanja, posebno ukoliko gledamo specifične sredine (recimo gradove). Također, u periodu od 2013. godine svjedoci smo trenda iseljavanja mladog stanovništva iz cijele BiH, ali posebno iz regija Krajine i Hercegovine. Struktura stanovništva tih predjela se sigurno značajno promijenila u periodu od 8-9 godina. Čak i bez ovih trendova specifičnih za strukturu stanovništva u BiH, generalni trendovi idu ka povećanju visokobrazovanog stanovništva i općenito produžavanju vremena provedenog u obrazovanju, što znači da se sa protokom vremena struktura stanovništva u tom segmentu stalno mijenja.

Razlog zbog kojeg ovaj uzorak svrstavamo u neprobabilističke se ogleda u tome što jedinke unutar kvota ne biramo nekim znanstvenim postupkom nego najčešće po dostupnosti. To u praksi znači da istraživači na terenu imaju uputu da prikupe ispitanike sa određenom karakteristikom u određenom broju, i iako postoji implicitna tendencija da se u okviru tog odabira osigura reprezentativnost, ona nije ničim osigurana, te istraživači mogu birati ispitanike koji su im iz nekog razloga dostupniji čime će uzorak učiniti pristrasnim. Dostupnost se može ogledati u činjenici da se ustvari biraju dobrovoljci (za koje smo već rekli da postoje saznanja da se značajno razlikuju od ostatka populacije, što je za korelacijska istraživanja gdje nam je izuzetno bitna vanjska valjanost veliki problem), ali i u činjenici da su npr. stanovnici određenog područja istraživačima dostupniji pa se kvote više “pune” njima.

Iz gore navedenih razloga, za kvotni uzorak se može reći da daje privid reprezentativnosti postavljanjem i ispunjavanjem kvota. Problem je u tome “šta” je u kvotama, tj, ne samo u načinu definiranju kvota nego i u odabiru jedinki u unutar kvota. Kako bi se premostio ovaj nedostatak anketarima se nekada unaprijed odredi područje sa kojeg moraju anketirane osobe biti, čime se pokušava smanjiti pristrasnost.

S druge strane, kvotni uzorak je izuzetno jednostavan i ekonomičan za primjenu. Materijalni troškovi primjene su manji nego kod primjene probabilističkih uzoraka, pitanje definiranja okvira uzorkovanja je skoro nepostojeće, a istraživačima, tj. anketarima su date jednostavne upute koje mogu relativno lagano pratiti.

Namjerni uzorak je vrsta neprobabilističkog uzorka koji se koristi kada koristimo neka postojeća saznanja o populaciji i pravimo slaganje sa našim istraživačkim ciljevima. U ovom slučaju odabir uzorka se pravi na osnovu istraživačeve procjene jedinki koje će najbolje odgovoriti cilju istraživanja. Ovakav pristup je ekonomičan iz razloga što značajno smanjuje potrebno vrijeme za prikupljanje uzorka, kao i materijalne troškove prilikom odabira uzorka.

Vrlo često u ovaj uzorak biramo tkz. tipične slučajeve tj. tipične pripadnike određene skupine koja nam je od interesa. Potencijalni problem sa ovakvim načinom uzorkovanja je na prvu očit. Radi se o načinu odabira tipičnog slučaja u kojem ukoliko dođe do greške imamo problem sa komplet valjanosti istraživanja. Verzija namjernog uzorka je i uzorkovanje kritičnog slučaja, tj, jedinki koje imaju karakteristike koje su posebno relevantne ili informativne za ilustraciju određenog problema. Ovakvo uzorkovanje se primjenjuje prilikom prognoziranja pobjednika izbora u SAD-u, gdje važi princip da su određene države kritične za konačnog pobjednika iz razloga što nisu tipično republikanske ili tipično demokratske, pa pobjeda u tim državama često odnosi prevagu.

Primjena tehnike snježne grude podrazumijeva način uzorkovanja u kojem istraživač odabere jedinke koje imaju odgovarajuće karakteristike neophodne za sudjelovanje u istraživanju, a onda uz pomoć tih jedinki dolazi do većeg broja pojedinaca istih karakteristika. Ovakav postupak je izuzetno koristan kada su nam za uzorak potrebni ispitanici vrlo specifičnog profila do kojih je teško doći. Pristrasnost koja se javlja kod ovakvog načina uzorkovanja je opet vezana za faktor dostupnosti gdje su nam dostupniji komunikativniji, društveniji i poznatiji pojedinci (što ih čini odvojenim podskupom unutar populacije).

Dakle, neprobabilistički uzorci su značajno lakši za primjenu i ekonomičniji. Primjenjujemo ih kada nemamo fokus na vanjskoj valjanosti, tj. kada ne postoji isključivi imperativ reprezentativnosti.

II) Probabilistički način uzorkovanja, tj. probabilističke uzorke karakterizira postojanje poznate vjerovatnoće za svakog pripadnika populacije da uđe u uzorak. Pripadnici populacije nemaju nužno jednaku vjerovatnoću da uđu u uzorak, ali činjenica da je ta vjerovatnoća poznata znači da je moguće napraviti prilagodbe kako bi se ispravila situacija za jedinke kod kojih je ta vjerovatnoća manja. Probabilističko uzorkovanje nužno podrazumijeva primjenu postupka slučajnog izbora ispitanika u istraživanje. Primjenom metode slučaja otklanjaju se opasnosti pristrasnosti i subjektivnosti istraživača čime se omogućava primjena statističkih postupaka za procjenu parametara populacije na osnovu uzorka.

Primjena metode slučajnog odabira je sistematičan postupak koji osigurava da je odabir svake naredne jedinice u uzorku neovisan od prethodnog odabira. Slučajni postupak je u osnovi svake vrste probabilističkih uzoraka, i s obzirom na to se svi ostali postupci koji se primjenjuju smatraju podvarijantama koje trebaju osigurati praktičnost i ekonomičnost implementacije.

Jednostavni slučajni uzorak je onaj uzorak u kojem je osigurano da svi pripadnici ciljane populacije imaju jednaku vjerovatnoću da uđu u uzorak. Dakle, primjena jednostavnog slučajnog uzorka podrazumijeva da imamo ili spisak svih pripadnika populacije, ili da su svi pripadnici populacije prisutni prilikom uzorkovanja na istom mjestu. Nakon toga, iz populacije se biraju pojedinci za uzorak primjenom neke od tehnika slučajnog odabira dok se ne ispuni zacrtani broj jedinki u uzorku. U ovom slučaju svi pripadnici populacije imaju jednaku vjerovatnoću da uđu u uzorak, koja je jednaka n/N (gdje je n - broj ispitanika u uzorku, N - broj jedinki u populaciji).

Sa statističke strane primjena jednostavnog slučajnog uzorka ima brojne prednosti. Većina softverskih statističkih postupaka ima ugrađene procjene populacijskih parametara koji počivaju na primjeni slučajnog uzorkovanja. Nedostaci primjene ovog postupka uzorkovanja su vezani za praktičnost i ekonomičnost. Kompletan spisak svih pripadnika populacije nam je rijetko dostupan. Primjena slučajnog odabira ne uzima nikakve praktične parametre u obzir pa su mogući veliki materijalni troškovi za primjenu istraživanja (recimo da su svi ispitanici iz različitih gradova i sl)

Slučajni sistemski uzorak je izuzetno sličan jednostavnom slučajnom. Razlika se ogleda u tome da se donese odluka da se u uzorak uvrštava svaki i -ti ispitanik sa spiska gdje se je i konstanta koja iznosi $i=N/n$ gdje je N broj ispitanika u populaciji a n - planirani broj ispitanika u populaciji. Ovdje je potrebno napomenuti da se ispitanik (tj. redni broj na spisku populacije) od kojeg ćemo krenuti također mora odabrati slučajnom metodom (a ne krenuti od početka i sl.). U slučaju u kojem bismo odabrali prvog sa spiska za prvog sudionika uzorka, ostali ispitanici koji ulaze u uzorak su onda unaprijed određeni i vjerovatnoća za ostatak populacije da uđe u uzorak je nula, što ovaj uzorak onda čini neprobabilističkim. Prednost ovakvog uzorkovanja je i u tome što ne moramo imati spisak svih pripadnika populacije. Možemo recimo u uzorak uvrštavati svakog i -tog kupca neke prodavnice, ili posjetioca neke institucije i sl. Problem sa primjenom ove tehnike uzorkovanja se ogleda u tome da redosljed jedinica populacije na spisku nije uvijek slučajan (jedinke koje imaju određenu karakteristiku se često na spisku grupiraju zajedno, kao što su recimo stanovnici nekog naselja ili osobe određene struke). Redosljed koji bi bio odgovarajući je recimo abecedni po prezimenu ili imenu.

Stratificirani uzorak zahtjeva da je prije početka samog uzorkovanja populacija razdjeljena na stratumе tj. podskupine. Članovi populacije se s obzirom na određenu karakteristiku svrstavaju u jedan stratum, nakon čega se iz svakog stratuma metodom slučaja biraju ispitanici za uzorak. Stratifikacija ima smisla jedino ukoliko je mjerena osobina povezana sa određenim stratumom. Npr. ocjena prednosti softverskih aplikacija u obrazovanju je povezana sa ulogom koju ispitanik ima u obrazovnom procesu (učenik, nastavnik, posmatrač) pa populaciju od interesa ima smisla razvrstati u stratumе s obzirom na ovu ulogu. Ovaj korak je naročito važan, jer bez prethodnih saznanja o vezi između grupirajuće varijable i mjerene varijable, stratificiranje nema smisla. Prednost stratificiranog uzorka se ogleda u tome da se smanjuje vrijednost greške time što se smanjuje homogenost uzorka. Odličan primjer stratuma kod evaluacije lakoće korištenja i razumljivosti softverskih aplikacija je prethodno računarsko iskustvo. Ako ispitanike biramo isključivo među studentima informatike velika je mogućnost pogrešnih nalaza. Stratificirani uzorak može biti srazmjernan ili nesrazmjernan u zavisnosti od toga da li je prisutna razmjerna zastupljenost pripadnika stratuma u uzorku. Ukoliko koristimo nesrazmjernu stratifikaciju potrebno je rezultate naknadno ponderirati kako bi se udjeli u uzorku uskladili sa udjelima u populaciji.

Klaster uzorak je vrsta probabilističkog uzorka u kojem jedinice uzorkovanja nisu pojedinci nego skupine, klasteri. Primjer uzorkovanja klastera je istraživanje učenika u školama gdje se umjesto pojedinaca uzorkuju cijeli razredi. Osnovna prednost ovakvog načina uzorkovanja je u praktičnosti i manjim troškovima vremena i novca. Posebno nam je ovaj način uzorkovanja koristan kada nemamo adekvatan popis svih jedinica populacije.

Pored navedenih vrsta probabilističkih uzoraka imamo i **stupnjeviti i zonski uzorak**. Stupnjeviti uzorak je svojevrsna prelazna forma između jednostavnog slučajnog i uzorka klastera koji primjenjujemo kako bismo iskoristili prednosti i jednog i drugog načina uzorkovanja. Sam postupak uzorkovanja je kao i kod uzorka klastera sa modifikacijom da se osnovni postupci prilikom odabira klastera ponavljaju dva ili više puta do konačnog svrstavanja jedinice u uzorak. Kod zonskog uzorkovanja osnova za uzorkovanje nije spisak pripadnika populacije, nego geografsko područje. Kod ovog uzorkovanja koristimo geografske karte kao granice klastera.

U anketnim istraživanjima greškom posmatramo razliku između populacijskih parametara (stvarnih) i onih izmjerenih na našem uzorku. U tom smislu razlikujemo pristrasnost kao sistemsku grešku kojom je opterećen uzorak i promjenjivu grešku koja se smatra slučajnom.

3.4.4. Određivanje potrebne veličine uzorka

Veličina uzorka je direktno povezana sa smanjenjem ukupne greške. Ovdje je potrebno napomenuti da je povećanje preciznosti koje dobijemo sa povećanjem broja ispitanika veće za male uzorke, dok nakon određene veličine dobitak na preciznosti povećanjem broja ispitanika postaje neznačajan ili čak nepostojeći.

Odluka o veličini uzorka je ustvari proces vaganja koristi u smislu smanjenja greške i povećanja preciznosti mjerenja i povećanja troškova sprovođenja istraživanja (kako materijalnih tako i utroška vremena. Također, kao što je već obrazloženo, povećanje veličine uzorka nam može dati lažnu sigurnost u rezultate, jer veliki uzorak nužno ne znači i smanjenje greške. Generalno gledano, kada govorimo o anketnom istraživanju, tačnost rezultata tj. procjena u većoj mjeri zavisi o metodološkoj korektnosti planiranog nacrtu i metode uzorkovanja nego od same veličine uzorka.

3.5. Upitnici za mjerenje upotrebljivosti

Fajgelj (2005) navodi kao je upitnik mjerni instrument koji je sastavljen od stavki ili pitanja koja se ispitaniku prezentiraju odštampana na papiru ili na nekom drugom mediju (recimo računaru). Upitnik u kontekstu ispitivanja upotrebljivosti je sličan intervjuu prenesenom na papir. Upitnik se konstruira za potrebe konkretnog istraživanja i analiza odgovora se radi na nivou pitanja, dakle ne računa se ukupni skor.

Upitnici mogu sadržavati različite vrste stavki kao što su recimo dihotomne (da/ne pitanja), otvorene („Opišite svoje iskustvo korištenja softvera“), izbora između jedne od više mogućnosti (nikad, rijetko ili ponekad opcije) što uključuje i ocjenu stepena slaganja i pitanja otvorenog tipa. Odluka o tome koju vrstu pitanja koristimo ovisi od mnogo faktora. Obično u slučajevima kada nam je važno doći do veće količine materijala koji možemo dubinski analizirati (i odgovori su nam relevantni jer su nam ih ponudili ispitanici koji su eksperti u tom području) onda je dobro koristiti pitanja otvorenog tipa jer ćemo dajući slobodu ispitanicima da iskazuju svoje razmišljanje i stavove dobiti značajnu količinu relevantnih informacija. S druge strane, kada želimo prikupiti podatke od velikog broja ispitanika (recimo kada je proizvod koji testiramo nešto što će se koristiti u širokoj potrošnji) onda je dobro imati

upitnike sa pitanjima zatvorenog tipa kako bismo mogli analizirati veliku količinu podataka kroz statističke parametre.

Upitnici sa otvorenim pitanjima se koriste u situacijama kada želimo kompleksne odgovore ili percepciju ispitanika o kompleksnim stvarima, kada smo u početnoj fazi razvoja proizvoda i zanimaju nas sva korisnička zapažanja i sl. Problem sa administriranjem ovih upitnika je u tome što primjena može trajati dugo, ispitanici mogu biti nemotivirani da daju iscrpne odgovore i odbijene odgovore je teško bodovati i interpretirati.

Upitnici sa ponuđenim odgovorima obično sadrže dihotomne opcije ili politomne stavke kao što su Likertove skale procjene gdje se npr. stepen slaganja sa tvrdnjom iskazuje npr. na skali od 1- 5, ili brojčane vrijednosti prate i tekstualne kvalifikacije.

Prilikom dizajniranja upitnika potrebno je razmišljati o konkretnom cilju istraživanja, ciljnoj populaciji, načinu administriranja upitnika i sl. Nakon definiranja cilja upitnika potrebno se odlučiti za teorijsku osnovu na kojoj ćemo tražiti odgovore na naša pitanja i sa obzirom na koju ćemo interpretirati rezultate. Ovaj korak je izuzetno važan jer daje svojevrsno znanstveno utemeljenje onome što radimo, a istovremeno nam je kompas prilikom dizajna instrumenta i tumačenja rezultata. Teorijsko utemeljenje je važno i kako bismo osigurali da konkretan problem koji nas zanima obuhvatimo u potpunosti, sa svih relevantnih aspekata. Praksa, svojevrsna grupna promišljanja (eng. *brainstorming*) prilikom dizajniranja instrumenta, dakle nabacivanje ideja u momentu i formuliranje čestica, korektna je jedino u slučaju kada je to pretkorak nakon kojeg slijedi proces pilot istraživanja u svrhu validacije instrumenta.

Generalno, u svakom istraživanju u kojem pravimo instrument od početka pilot istraživanje je poželjan korak. Pilot studija podrazumijeva da napravimo nešto obimniji instrument (više čestica nego što planiramo imati u konačnoj verziji) i primijenimo taj instrument na većem uzorku ispitanika. Analizom odgovora na proširenu listu pitanja, kao i analizom komentara ispitanika radimo selekciju čestica koje će ostati u konačnoj verziji. Analiziramo odgovore ispitanika na svakoj čestici. Rezultati koji su kontradiktorni nam ukazuju na to da ispitanici neku od ponuđenih čestica nisu razumjeli ispravno, tj. da naša formulacija nije adekvatna. Ponovnom analizom sadržaja pokušavamo identificirati tu česticu kako bismo je odstranili. Također, čestice na kojima imamo uglavnom iste odgovore nisu diskriminativne te ih iz tog razloga eliminiramo. Analiziramo i komentare ispitanika koji su popunili pilot instrument i pokušavamo otkloniti ostale probleme na koje su nam ukazali. U ovoj fazi je i dobro imati svojevrsnu kontrolnu skupinu profesionalaca koji imaju iskustva u istraživanju fenomena od našeg interesa. Od profesionalaca prikupljamo kvalitativnu povratnu informaciju koju koristimo za unapređenje instrumenta. U pilot verziji je idealno trostruki broj pitanja od onog koji želimo u konačnoj verziji instrumenta. Minimum je 1/3 više pitanja. Dakle, ukoliko odaberemo adekvatno teorijsko utemeljenje za dizajn instrumenta, te pitanja odaberemo na osnovu već postojećih iz dostupnih instrumenata možemo u određenoj mjeri odstupiti od gore opisanog rigoroznog procesa prethodne evaluacije instrumenta, ali je svakako poželjna prethodna provjera instrumenta u nekom obliku prije nego ga primijenimo kao finalnu verziju u istraživanju.

3.5.1. Dizajniranje anketnog upitnika

Cilj dizajna anketnog upitnika, tj. formuliranja pitanja sadržanih je potaknuti ispitanike da izkažu svoje mišljenje ili stav glede nekog pitanja. Anketne upitnike ne karakteriziraju dobre metrijske karakteristike kao ostale psihološke instrumente. Dodatni problem se ogleda u činjenici da nam često metrijske karakteristike anketnih upitnika nisu ni poznate. Anketnim upitnikom mi ne pokušavamo otkriti objektivno stanje stvari, nego želimo doći do subjektivnih mišljenja i stavova ispitanika (dakle, mi anketom ne možemo doći do informacije o stanju voskog obrazovanja u BiH, ali možemo prikupiti mišljenja studenata, nastavnika i poslodavaca što nam može pomoći da donesemo zaključke o stvarnom stanju stvari).

Prilikom definiranja pitanja potrebno je posebnu pažnju posvetiti odluci o tome da li ćemo primijeniti otvorena ili zatvorena pitanja. Otvorena pitanja ne nude unaprijed osmišljene odgovore između kojih se ispitanik odlučuje kao što je to slučaj kod pitanja zatvorenog tipa.

Pitanje: “Kako biste opisali vaše digitalne kompetencije?”, je pitanje otvorenog tipa gdje od ispitanika očekujemo da sami ponude odgovor, bez sugeriranja mogućih tj. ponuđenih odgovora.

Primjer zatvorenog pitanja bi bilo formuliranje na sljedeći način:

“Kako biste ocijenili vaše digitalne kompetencije?”

- a) Profesionalni nivo stručnosti - koristite efektivno, sveobuhvatno i fleksibilno široki repertoar digitalnih strategija
- b) Napredni korisnik - koristite vješto i kreativno veliki broj digitalnih alata i aplikacija kako biste unaprijedili svoju praksu
- c) Prosječan korisnik - koristite i eksperimentirate sa digitalnim alatima, pokušavate na naučiti kako da ih primijenite u kontekstu svog rada
- d) Bazno poznavanje - koristite digitalne alate, ali nemate konzistentan i sveobuhvatan uvid, pa vam treba inspiracija od drugih”

Dakle, kod pitanja zatvorenog tipa mi ispitanicima nudimo određene opcije odgovora i oni mogu birati između tih opcija. Zatvorena pitanja imaju prednosti koje se ogledaju u lakšoj i bržoj primjeni, lakšoj uporedbi ispitanika s obzirom na to da su svi odgovori raspoređeni u kategorije. Posebno je značajna ušteda vremena prilikom obrade podataka dobijenih zatvorenim pitanjima, zbog toga što je moguće korištenje računarskih alata od samog početka do kraja obrade.

Davanje odgovora na otvorena pitanja od ispitanika zahtjeva više napora i vremena što često dovodi do toga da ispitanici odbiju odgovoriti na pitanje. Postupak analize odgovora dobijenih otvorenim pitanjima je također puno teži i dugotrajniji jer zahtjeva razvrstavanje odgovora po kategorijama pa smještanje pojedinačnih odgovora u kategorije, i onda tek postoji opcija računarske obrade. Kod zatvorenih pitanja moguće je da ne predvidimo i samim tim ne ponudimo odgovor koji bi ispitanik izabrao.

Zatvorena pitanja su najčešće tipa višestrukog izbora i skale ocjenjivanja. Skale ocjenjivanja obično od učesnika traže da ocijene apstraktne i subjektivne koncepte, kao što

su zadovoljstvo, slaganje ili neslaganje sa tvrdnjama, itd. Ponuđene tvrdnje i odgovori kod zatvorenih pitanja mogu imati veliki utjecaj i na odgovore ispitanika, a samim tim i na interpretaciju. U nastavku ćemo navesti neke primjere formuliranja zatvorenih pitanja u anketnom upitniku adaptiranih iz Saurovog pregleda različitih mjernih skala (Sauro 2018).

Linearna numerička skala ocjenjivanja je skala na kojoj ispitanici daju numerički odgovor na pitanje ili numeričku ocjenu izjave. Pitanja se odnose na zadovoljstvo ili lakoću korištenja, važnost neke osobine ili vjerovatnoću daljeg korištenja, kupovine ili preporuke drugima. Primjer pitanja je dat na slici 3.1. Uobičajeno je da se krajnjim numeričkim vrijednostima pridruže i tekstualne oznake.

Ocijenite koliko je vjerojatno da ćete ovu aplikaciju preporučiti prijateljima?

1 2 3 4 5 6 7 8 9 10

Neću preporučiti Sigurno ću preporučiti mnogima

Slika 3.1. Primjer linearne numeričke skale ocjenjivanja

Likertova skala slaganja je skala koja se koristi za mjerenje mišljenja i stavova. Koristi niz odgovora u rasponu od jednog ekstremnog stava do drugog. Budući da je skala odgovora o slaganju, stavke trebaju biti takve da se učesnici mogu složiti ili ne složiti. Klasična Likertova stavka koristi skalu odgovora od 5 tačaka, ali se može koristiti i 7, 9 ili drugo. Primjer sa pet tačaka kojima su pridružene tekstualne oznake je dat na slici 3.2.

Uopšte se ne slažem Ne slažem se Ne mogu odgovoriti Slažem se Potpuno se slažem

Upute su mi pomogle da pronađem izlaz iz dvorišta.

Slika 3.2. Primjer Likertove skale slaganja

Ovaj primjer sa neparnim brojem tačaka uključuje u sredini skale neutralnu tačku. Moguće je imati i parni broj tačaka, što znači da se korisnik mora odlučiti za izbor između slaganja i neslaganja. U nekim kontekstima ovo može biti opravdano, ali kada je u pitanju subjektivno korisničko iskustvo, poželjno je korisniku ostaviti mogućnost da je neutralan.

Takođe, treba voditi računa da korisnik iz nekog opravdanog razloga, nema stav po pitanju navedene tvrdnje i tada neutralni odgovor je jedini tačan. Kod upotrebljivosti svaki vid nametanje odgovora se tumači kao metrika taštine, odnosno „pecanje“ pohvala od ispitanika. Kada su brojevi pridruženi svakoj opciji odgovora, Likertova stavka se može posmatrati kao poseban slučaj linearne numeričke skale.

Matrica ocjenjivanja više tvrdnji je kompaktan način predstavljanja višestrukih linearnih numeričkih skala, uključujući i Likertove skale slaganja. Na ovaj način se korisniku olakšava odgovor. Isječak iz jednog upitnika sa primjerom matrice ocjenjivanja stavova je dat na slici 3.3.

Zamoliti ćemo Vas da ocijenite koliko ste saglasni sa tvrdnjama o značaju i mogućnosti korištenja obrazovnih sadržaja tipa Coursera kurseva u nastavi *

	Ne slažem se	Djelimično se ne slažem	Nemam stav	Djelimično se slažem	Slažem se
Obrazovni sadržaji koje nudi Coursera su dobro osmišljeni	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Obrazovni sadržaji koje nudi Coursera su usklađeni sa najnovijim spoznajama.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Coursera nudi dosta kurseva koji tematski odgovaraju onom što ja podučavam	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Slika 3.3. Primjer matrice ocjenjivanja više tvrdnji Likertovom skalom slaganja

Skala učestalosti pojava je specifična skala slaganja usmjerena ka tvrdnjama vezanih za učestanost korištenja. Uz skale slaganja sa osobinama proizvoda, uvid u korisničko zadovoljstvo se može dobiti i mjerenjem učestanosti korištenja ili namjere korištenja. Kao odgovori na pitanja koliko često se koristi neka aplikacija nude se numeričke odrednice, ali i općeniti vremenski okviri: ponekad, uvijek, nikad – što se naziva verbalna skala učestalosti. Bitno je da su ponuđeni odgovori sa frekvencijama poredani i da su precizni.

Skala poređenja je skala kod koje korisnici mogu da ocijene svoj izbor između dvije alternative, i to da iskažu i stepen svoje preferencije koristeći skalu poređenja. Na slici 3.4. je ilustrovan primjer kada se od ispitanika traži da ocijene svoje preferencije za dvije različite aplikacije Skok sa starog mosta – web i VR.

Nametnuto rangiranje predstavlja skale koje se koriste za određivanje prioriteta kod evaluacije softvera i namjera korištenja. Ako se korisnicima prepusti da slobodno ocjenjuju svoje interese i namjere na linearnoj numeričkoj skali, može doći do problema da korisnici sve ili većinu karakteristika ocijene visoko i da se ne može izvršiti rangiranje. Zato je korisno da se kod dizajniranja upitnika mogu koristiti pitanja sa nametnutim rangiranjem. Načini implementacije uključuju drag-and-drop pristup kada se stavke prevlače iz jedne liste u drugi i slažu po prioritetu, ili ako nema ovakve vizualne kontrole na interfejsu, potrebno je

ponuditi numeričku skalu jednake dužine kao i lista stavki, pa da korisnici izvrše uparivanje stavki i numeričkih rangova.

Ponavljamo iste izjave kao i dijelu II ali sad se trebete opredijeliti tako što birate tip aplikacije za koju smatrate da je tvrdnja više tačna. Ako mislite da je tvrdnja netačna za obje ili se ne možete odlučiti kojoj aplikaciji da date prednost izaberite odgovor u sredini: "Jednake su."

Uporedite VR aplikaciju Skok sa Starog mosta i web aplikaciju postavljajući vaš *
odgovor na sljedećoj skali:

	Web - izuzetno	Web	Web - umjereno	Jednake su	VR - umjereno	VR	VR - izuzetno
Korištenje aplikacije pruža osjećaj zaigranosti i maštovitosti.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aplikacija mi je omogućila stvarno iskustvo visine.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aplikacija mi je omogućila da razumijem iskustvo skakanja sa Starog mosta.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Slika 3.4. Primjer skale poređenja

3.5.2. Izvori greške vezani za anketni upitnik

I anketne upitnike, kao i ostale mjerne instrumente u društvenim istraživanjima moramo promatrati kroz postojanje dva vrlo važna svojstva: valjanosti i pouzdanosti. Valjanost ankete nam daje odgovor na pitanje da li anketom stvarno mjerimo ono što mislimo da mjerimo. U anketnim istraživanjima valjanost ugrožava davanje društveno poželjnih odgovora od strane ispitanika, strah ili nepovjerenje.

Pouzdanost ankete se odnosi na dosljednost ispitanikovih odgovora pri istovjetnom i sličnom pitanju kroz vrijeme.

Do smanjenja pouzdanosti dolazi zbog načina postavljanja pitanja, redoslijeda pitanja, neposjedovanja stava o temi o kojoj se ispitanici pitaju ili davanja društveno poželjenih odgovora.

3.5.2.1. Izvori greške vezani za način postavljanja pitanja

Više pitanja u jednom je česta greška koju istraživači prave prilikom formiranja anketnog upitnika. Najčešće se u jednom pitanju traži stav ili mišljenje o više stvari, gdje je moguće da se ispitanik slaže sa jednim dijelom, ali ne i sa drugim. U ovim slučajevima ispitanik najčešće odbija odgovoriti ili nasumice odgovori ustvari na samo jedan dio pitanja (onaj s kojim se slaže ili onaj s kojim se ne slaže), gdje istraživač ustvari ne dobija uopće pravu informaciju. Primjer ovakvog pitanja bi bio: “Smatram da je potrebno dići porez na cigarete kako bi se finasirala izgradnja autoputa.” Ispitanik može recimo smatrati da je potrebno povećati porez na cigarete, ali bi imao sasvim druge ideje šta učiniti sa tim novcem i generalno se ne slaže sa takvim načinom finansiranja izgradnje autoputa.

Također, greška koja se često pravi prilikom formulacije pitanja se odnosi na višeznačnost ili nejasnost formulacije. Ovakva pitanja sadrže fraze kao što je: “najčešće”, “rijetko”, “obično” i sl. Problem sa upotrebom ovih odrednica je u tome što ih ispitanici različito shvataju. “Obično” može značiti potpuno različitu frekvenciju kod različitih ispitanika. Pored toga, pitanje “Obično ujutro vježbam” ispitanici mogu shvatiti na više načina (1. kada vježbam, vježbam ujutro prije nego drugim dijelovima dana ili 2) moje jutarnje aktivnosti obično podrazumijevaju da vježbam). Kako bi se ove greške izbjegle potrebno je precizno definirati tj. operacionalizirati na šta tačno mislimo npr. umesto “često” navedemo pet i više puta sedmično i sl.

Prilikom formuliranja pitanja bitno je da razmišljamo o tome da ih formuliramo uravnoteženo. Npr. pitanje formulirano na način: “Smatrate li da VR aplikacija skokova sa Starog mosta pruža bolje korisničko iskustvo u donosu na web aplikaciju?” je neuravnoteženo i kod nekih ispitanika može dovesti do veće vjerovatnoće davanja potvrdnog odgovora nego da smo pitanje formulirali na sljedeći način: “Koja verzija aplikacije skokova sa Starog mosta pruža bolje korisničko iskustvo VR ili web?”.

Odabir riječi je izuzetno bitan prilikom formulacije pitanja. Potrebno je birati riječi koje će nam omogućiti da sažeto i jasno formuliramo ono što želimo pitati. Pri tome treba voditi i računa o stepenu razumijevanja ispitanika, dakle ne treba koristiti stručne termine ili generalno vokabular neprimjeren nivou obrazovanja ispitanika.

Pitanja ne smiju biti sugestivna, tj. ne smiju biti takva da navode ispitanika na određeni odgovor. Ukoliko formuliramo pitanje na način: “Mnogi kažu da je porodica i porodične vrijednosti u krizi, da li se slažete s tim?”, mi ustvari ispitanika navodimo na očekivani odgovor. U ovakvoj formulaciji mi već iznosimo tvrdnju (koja možda ispitaniku ne bi uopće bila u svijesti na ovakav način formulisana) i pri tome dajemo informaciju o tome da “mnogi” misle tako. Puno bolja alternativa bi bilo pitanje da nam ispitanici kažu šta misle o stanju porodice i porodičnih vrijednosti.

Pristup dizajniranju otvorenih i zatvorenih pitanja se razlikuje, i svaki od ovih tipova pitanja imaju svoje izazove prilikom implementacije. Nedostatak zatvorenih pitanja se ogleda u činjenici da će ispitanici vjerovatnije odabrati neki odgovor od ponuđenih čak i u slučaju kada nemaju formirano mišljenje ili informaciju. S druge strane, kada postavimo pitanje otvorenog tipa manja je vjerovatnoća za tako nešto. Također, pitanje formuliranja alternativa

i određivanje broja alternativa je vrlo osjetljivo. Ukoliko ponudimo mali broj alternativa postoji mogućnost da ispitanik u ponuđenim opcijama ne pronađe onu koja mu odgovara. Prevelik broj alternativa je zamoran ispitanicima za čitati i gubi se na osnovnoj prednosti primjene ovih pitanja a to je ekonomičnost.

Kako bi se maksimalno iskoristile prednosti obje vrste pitanja, te kontrolisali nedostaci najbolje je u različitim fazama istraživanja koristiti različite vrste pitanja. Pitanja otvorenog tipa je najbolje koristiti u pilot istraživanju kako bi se skupilo što više različitih odgovora. Analizom odgovora prikupljenih u pilot istraživanju možemo generirati kvalitetnije alternative koje ćemo ponuditi u finalnoj verziji upitnika. Finalnu verziju primjenjujemo na većem uzorku gdje nam je ekonomičnost i praktičnost primjene i analize važna.

Poseban problem kod implementacije anketnog istraživanja je situacija kada nas zanima mišljenje ispitanika o temi koja nije pretjerano aktualna u općoj populaciji. Često nas zanimaju pitanja od značaja uglavnom akademskoj zajednici ili određenoj interesnoj skupini (npr. profesionalcima određenog područja) dok ostatak populacije nema formirano mišljenje niti informacije o temi. Iz tog razloga važno je ponuditi alternativu “ne znam” ili neki ekvivalent toga. Bez ove alternative ispitanici koji nemaju mišljenje o onome što ih pitamo će nasumično zaokrižiti jednu alternativu što svakako ne želimo.

Redoslijed pitanja je varijabla o kojoj često ne razmišljamo prilikom dizajniranja anketnog upitnika. Isto pitanje u zavisnosti od toga koja mu pitanja prethode je može biti potpuno različit podražaj. Kako bi se izbjegle prijetnje valjanosti ovog tipa, pitanja je potrebno složiti logičkim redom na način da su organizirani s obzirom na teme i cjeline, ali i na vrstu pitanja (otvoreni i zatvoreni tip, da-ne pitanja ili pitanja višestrukog izbora). Generalno, pitanja jednog tipa grupiramo zajedno, redamo ih logičkim slijedom po temama. Počinjemo sa pitanjima da-ne, prelazimo na pitanja višestrukog izbora, pitanja u kojima ispitanici trebaju dopuniti kratki odgovor (riječ ili dvije) i onda na kraju ostavljamo pitanja otvorenog tipa. Navedeni redoslijed se naravno odnosi na ankete koje sadrže sve ove vrste pitanja, što nije neophodno prilikom dizajniranja ankete.

3.5.2.2. Društveno poželjni odgovori

Poseban problem s kojim se suočavamo prilikom ovakog načina ispitivanja mišljenja, stavova ili ponašanja predstavlja davanje društveno poželjnih odgovora. Ljudi su često neskloni da prijave mišljenja za koja smatraju da odstupaju od većine ili od onoga što smatraju da je generalno društveno poželjno. Milas (2005) navodi da su ljudi nevoljni prijaviti kupovanje jeftinije odjeće ili da ne idu u pozorište i sl. banalne stvari. Kako bi se izbjegla ili ako ništa smanjila vjerovatnoća davanja društveno poželjnih odgovora, prvi korak koji trebamo učiniti je osigurati anonimnost ispitanika.

Dalje, neophodno je da pazimo kako formuliramo pitanja i sa aspekta davanja “tragova” o tome šta je očekivano, “odobreno” ili “poželjno”. Pitanja trebaju biti formulirana na način da je ispitanicima jasno da je svaki odgovor koji je iskren, samim tim što je iskren i adekvatan. Dobar postupak je i također formulirati pitanje na način da se poricanje ponašanja predstavi malo vjerovatnim, tamo gdje je to stvarno malo vjerovatno. Npr. umjesto da pitamo ispitanike da li su nekada razmišljali o tome da su pogriješili u odabiru bračnog partnera, bolje i pitanje

formulirati na način da nas zanima koliko puta, ili koliko često sebi postavljaju pitanje ispravnosti odluke o odabiru partnera. Pored načina formuliranja pitanja, vrlo je važna i anketarova profesionalnost prilikom bilježenja odgovora. Anketar ni na koji način ne smije prilikom komunikacije sa ispitanikom ili posmatranja davanja odgovora pokazati reakcije odobravanja ili neodobravanja.

3.5.2.3. Sklonost potvrđivanju

Bez obzira na koji način i kakvo je pitanje postavljeno, postoji sklonost potvrđivanju od strane ispitanika. Sklonost potvrđivanju znači sklonost davanju odgovora "slažem se" ili nekog ekvivalenta tog odgovora. Ova sklonost značajno prijeti valjanosti generalno svih istraživanja. Ispitanici se dakle prilikom odgovaranja ne vode isključivo vlastitim mišljenjem nego dio odgovora daju automacki, gdje su skloniji da kažu da se sa tvrdnjom slažu, nego da se ne slažu. Milas (2005) navodi da su potvrđivanju, dakle davanju nasumičnih odgovora skloniji ispitanici koji su skomnijeg obrazovanja. Jedan od načina prevazilaženja ovog problema je formulacija pitanja na način da se od ispitanika ne traži da se izjasne o slaganju sa tvrdnjom nego da se odluče za jednu od ponuđenih tvrdnji. Navedeno je samo prividno rješenje jer ispitanici koji su skloni potvrđivanju će u situaciji u kojoj trebaju birati između alternativa davati nasumične odgovore.

3.6. Standardizirani upitnici za mjerenje upotrebljivosti

Standardizirani upitnik je upitnik dizajniran za ponovljenu upotrebu, obično sa određenim nizom pitanja predstavljenim u određenom redoslijedu u određenom formatu, sa određenim pravilima za izradu metrike na osnovu odgovora ispitanika. Kao dio razvoja standardiziranih upitnika, uobičajeno je da programer izvještava o mjerenjima njegove pouzdanosti, valjanosti i osjetljivosti - drugim riječima, da bi upitnik prošao psihometrijsku kvalifikaciju. Dakle, velika prednost standardiziranih upitnika je što su nam poznate njihove metrijske karakteristike, te su rezultati našeg istraživanja uporedivi sa rezultatima drugih istraživanja koja su koristila isti instrument te je rezultate koje smo dobili lakše iskomunicirati. Pored toga, razvijanje standardiziranih upitnika zahtjeva puno vremena, ali jednom razvijeni, vrlo su ekonomični za koristiti.

Sauro (2016) pravi pregled standardiziranih upitnika za procjenu zadovoljstva učesnika upotrebljivošću proizvoda. Najčešće korišteni upitnici za evaluaciju percepcije korisničkog iskustva nakon završetka zadatka ili nekog scenarija:

- Upitnik za zadovoljstvo korisničkom interakcijom (eng. *Questionnaire for User Interaction Satisfaction*) - QUIS (Chin i sur., 1988);
- Inventorij mjerenja upotrebljivosti softvera (eng. *Software Usability Measurement Inventory*) - SUMI (McSweeney, 1992; Kirakowski and Corbett, 1993);
- Post-studijski upitnik za upotrebljivost sistema (eng. *Post-Study System Usability Questionnaire*) - PSSUQ (Lewis, 1990a, 1992, 1995, 2002)
- Skala upotrebljivosti softvera (eng. *Software Usability Scale*) - SUS (Brooke, 1996).

- | | |
|------|---|
| P1. | Često bih koristio/koristila ovu aplikaciju |
| P2. | Smatram da je aplikacija nepotrebno kompleksna |
| P3. | Mislim da je aplikacija jednostavna za korištenje |
| P4. | Smatram da mi je potrebna pomoć tehničke osobe za korištenje aplikacije |
| P5. | Smatram da su različite funkcionalnosti dobro integrirane |
| P6. | Smatram da ima previše nedosljednosti u interfejsu aplikacije |
| P7. | Mislim da bi većina ljudi brzo naučila koristiti ovu aplikaciju |
| P8. | Mislim da je aplikacija veoma zahtjeva za korisnika |
| P9. | Osjećam se samopouzdana prilikom korištenja ove aplikacije |
| P10. | Neophodno je da naučim mnogo stvari prije korištenja ove aplikacije |

Slika 3.5. Primjer SUS upitnika od deset pitanja

Navedene skale se u praksi uobičajeno i uspješno primjenjuju prevedene na različite jezike, i prilagođene za različite vrste softverskih aplikacija i različite kontekste, i bez provedene validacije. Primjer je SUS skala za koju je detaljnija analiza primjene publikovana u knjizi (Sauro 2011) gdje autor navodi uspješnu primjenu nakon prevoda na španski, francuski i holandski, sa pokazanom sličnom pouzdanosti kao kod originalne engleske verzije. Na slici 3.5. je dat primjer SUS upitnika preveden na bosanski jezik. Ipak, u literaturi se sreću opisi stručnog prevoda i prilagodbe lokalnom kontekstu – jezik i validacija – za portugalski (Martins i sur. 2015) i malajski (Marzuki i sur. 2018), što otvara mogućnost istraživanjima sa ciljem validiranja ovih upitnika na uzorku stanovništva Bosne i Hercegovine, tako da se doprinese znanstvenoj težini primjene ovih instrumenata što bi značajno doprinijela snazi zaključaka do kojih dolazimo. Ovo je još jedan primjer moguće saradnje stručnjaka iz domena psihometrije i interakcije čovjek računar.

Često se razvijaju specifični instrumenti mjerenja fokusirani na pojedine aspekte upotrebljivosti kao što su kognitivno opterećenje, prihvatanje nove tehnologije, ili specifičnosti vezane za tip i namjenu aplikacija. Zbog značaja navedenih tema za upotrebu softvera u obrazovanju, knjiga uključuje posebna poglavlja koja se bave konceptima kognitivnog opterećenje i prihvatanja nove tehnologije, kao i instrumentima za njihovo mjerenje.

Kod softvera u obrazovanju izdvaja se specifičan skup aplikacija koje se nazivaju ozbiljne igre (eng. *Serious Games*). Uspjeh ovih aplikacija se uz mjerenje rezultata učenja mjeri konstruktima kao što su imerzija i edutainment. Edutainment je osobina izvedena kombinacijom obrazovne (eng. *education*) komponente i zabave (eng. *entertainment*) što je značajna osobina ozbiljnih igara.

3.7. Ostali načini mjerenja u istraživanjima upotrebljivosti

3.7.1. Vrijeme završetka zadatka

Ova mjera najčešće podrazumijeva vrijeme potrebno da se neka aktivnost/zadatak uspješno završi. Pored toga, nekada su nam potrebne informacije o tome koliko vremena korisnik provede koristeći aplikaciju prije nego odustane nakon neuspjeha kao i totalno vrijeme korištenja aplikacije (neuspješno i uspješno). Kada govorimo o obrazovnim softverima vrijeme završetka zadatka je posebno važna metrika. Vrijeme koje se vezuje za određenu aktivnost tokom učenja identifikovano je kao jedan od centralnih konstrukata koji utiču na uspjeh u učenju (Bloom 1974), a mjere vremena završetka zadatka se koriste za izgradnju prediktivnih modela učenja kako bi se razumjeli i poboljšali procesi učenja (Kovanovic i sur. 2016).

3.7.2. Kombinirana mjerenja

Različite metrike upotrebljivosti značajno koreliraju. Ipak, ne govorimo o toliko visokoj korelaciji koja bi opravdala zamjenu jedne metode drugom, tj. izjednačavanje tih postupaka. Generalno posmatrano, korisnici koji uspješnije izvršavaju zadatak, imaju i tendenciju da alat kojim to postižu ocjenjuju boljim i kvalitetnijim. Prikupljanje više metrika u testu upotrebljivosti je korisno jer na osnovu više izvora imao bolju sliku ukupnog korisničkog iskustva.

Poseban izazov predstavlja kombiniranje više metrika u jednu. U poslovnoj primjeni evaluacije korisničkog iskustva i upotrebljivosti postoje situacije koje zahtijevaju izbor jednog proizvoda na osnovu evaluacije upotrebljivosti za više proizvoda i pritom se trebaju osloniti na jedno mjerenje, tako da se ili opravda određivanje prioriteta zavisnih mjera, ili se koristi kompozitna mjera. I kad se oslanjamo na kompozitne zbirne podatke, važno je obezbijediti da podaci koji doprinose zbirnom rezultatu ostaju dostupni kao relevantna komponenta za sve analize i odluke koje zahtijevaju detaljnije informacije (Sauro 2016).

3.7.3. A/B testiranje

A/B testiranje je popularna metoda uporedbe dva softverska rješenja jedne aplikacije koja se obično razlikuju u jednoj karakteristici. Razlike u dizajnu dvije verzije aplikacije mogu biti i izuzetno male (npr. izgled jedne komande ili font koji se koristi u jednom dijelu). Podaci se prikupljaju na velikom uzorku i u tom smislu zadovoljavaju jako stroge metodološke kriterije.

Upotreba A/B testiranja je široko rasprostranjena pa je potrebno imati u vidu da ne postoji jedinstven način provođenja A/B testiranja - ali skoro svi pristupi A/B testiranju dijele određene zajedničke karakteristike. Svaki A/B test kreće od početne tvrdnje (hipoteze) koja povezuje uticaj promjene nezavisne varijable (način implementacije dijela interfejsa) sa modalitetima koji odgovaraju verzijama A i B na posmatranu zavisnu varijablu (neka metrike poslovanja i impakta naše aplikacije). Za A/B testiranje su potrebni kontrolni test (A) i alternativni test (B). Kontrolni test je u većini slučajeva može trenutna verzija aplikacije. A/B testovi se provode po eksperimentalnom nacrtu koji obezbjeđuje randomizirano uključivanje ispitanika, tako da se može utvrditi kauzalnost između nezavisne i zavisne varijable.

3.8. Zaključak

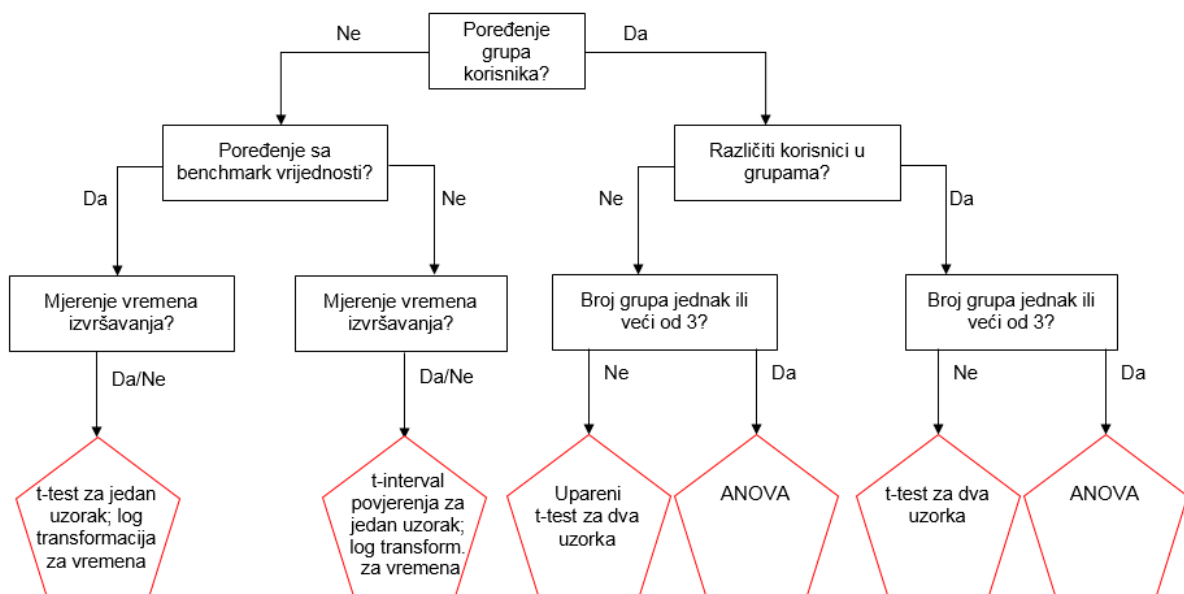
Opisani mjerni instrumenti su način prikupljanja podataka u studijama evaluacije korisničkog iskustva. U sljedećim poglavljima bit će opisan postupak dizajniranja eksperimentalnih i ne-eksperimentalnih studija, a zatim statistički postupci za analizu i interpretaciju prikupljenih podataka. Donošenje odluke o dizajniranju studije prvenstveno je vezano za tip podataka i skalu naših mjerenja. Kod tipova podataka razlikuju se pristupi analize kontinualnih i diskretnih podataka, posebno s obzirom na skale mjerenja koje su detaljnije opisane u početnom dijelu ovog poglavlja.

U nastavku na slici 3.6. nudimo primjer algoritma donošenja odluke o tome koji tip studije i koji nacrt koristiti, samo je ovaj dijagram algoritma donošenja odluke sada proširen u odnosu na sliku 1.2 i uključuje koji statistički postupci prate takvo istraživanje (Sauro 2016).

Kada nemamo pristup čitavoj populaciji korisnika, trebamo odluke donositi na osnovu mjerenja nad formiranim uzorkom. Veličina uzorka će uticati na preciznost naših mjerenja, a uz izvršenu analizu potrebno je da odredimo i preciznost dobijenih rezultata, odnosno koliko povjerenja možemo imati u dobijeni rezultat.

Prva odluka koju donosimo je vezana da li naša studija ima za cilj (1) procjenu mjere upotrebljivosti ili poređenje te mjere sa zadatim benchmarkom ili nam je cilj da (2) poredimo dvije ili više grupa po nekoj metrici upotrebljivosti.

Kod procjene preciznosti potrebno je da odredimo interval povjerenja oko izračunate metrike. U slučaju kada se radi sa metrikama vezanim za vrijeme izvršavanja, za koje se iskustveno zna da odstupaju od normalne raspodjele i da su nakrivljene, potrebno je izvršiti logaritamsku transformaciju prije primjene statističke analize preciznosti mjerenja.



Slika 3.6. Algoritam odlučivanja o studiji evaluacije upotrebljivosti

Kod poređenja dvije ili više grupa, pravi cilj je poređenje različitih aplikacija ili verzija aplikacije (kao kod A/B testiranja). U ovakvim studijama povjerenje u rezultate je veće ako isti korisnici testiraju različite aplikacije, ali tada imamo problem redoslijeda izloženosti različitim aplikacijama i stvaranja predrasuda, tako da se najčešće radi sa različitim korisnicima, što povlači upotrebu odgovarajućih statističkih testova. Često se analiziraju i razlike u ocjeni upotrebljivosti jedne aplikacije za ciljano različite grupe korisnika, npr. profesori i studenti, mađi i stariji korisnici itd.

Podsjećamo se da je cilj svih studija da se odredi da li je razlika mjerene vrijednosti u odnosu na benchmark ili razlika u metrici upotrebljivosti za dvije grupe korisnika statistički značajna ili je proizvod slučajnosti, odnosno varijacija u populaciji.

Studije evaluacije koje uključuju poređenje dvije ili više grupa korisnika omogućavaju analizu uzročnosti i smatraju se eksperimentima, s tim da kada uzorak nije formiran slučajnim odabirom smatramo ih kvazi-eksperimentima. Eksperimentalne i kvazi-eksperimentalne studije su detaljno pisane u poglavlju 4. Evaluacije koje ne uključuju analizu uzročnosti i vezuju se za mjerenje na samo jednoj grupi, se smatraju ne-eksperimentom, i opisane su u poglavlju 5. Kod studija tipa ne-eksperimenta posebno se izdvajaju studije kod kojih je fokus na mjerenju vremena izvršavanja zadataka. T-test koji se navodi na slici 3.6. će biti detaljnije objašnjen u poglavlju 6, a njegov značaj je u tome što uzima u obzir veličinu uzorka sa kojim raspoložemo u analizi.

4. Eksperimentalni nacrti studija upotrebljivosti

Prije nego ponudimo detaljniju analizu samog postupka provedena eksperimentalnog istraživanja, te statističkih postupaka koji se koriste u analizi rezultata ovakvih nacrti, neophodno je razjasniti jedno od osnovnih pitanja epistemologije – pitanje određivanja uzročnosti. Razlozi za elaboraciju pitanja dokazivanja uzročnosti su dvojaki. Prvo, zadatak svake znanosti je da predviđa pojave koje izučava. U slučaju psihologije, kao znanosti o ljudskom ponašanju, kao zadatak se nameće predviđanje ljudskog ponašanja. U istraživanjima upotrebljivosti implicitni cilj je predvidjeti koliko će korisnici koristiti proizvod te sa kakvim uspjehom. Kako bismo mogli predvidjeti ponašanje neophodno je da možemo identificirati uzroke ponašanja. Samo ukoliko možemo mapirati uzrok pojedinog ponašanja, onda je moguće dati procjenu o tome da li će se ponašanje pojaviti.

Drugi razlog za poseban osvrt na pitanje dokazivanja uzročnosti se odnosi na korištenje tog termina u svakodnevnom govoru i u medijskom prostoru. Vrlo često se u javnom prostoru prezentiraju rezultati istraživanja na način kao da je ustanovljena uzročno – posljedična veza između varijabli (npr. pametniji ljudi se više razvode, gdje se implicira da je viska inteligencija uzrokom razvoda braka).

Pitanje uzročnosti je posebno važno u kontekstu elaboracije eksperimentalnih nacrti, iz razloga što se upravo eksperimentalna metoda smatra metodom izbora za dokazivanje uzročnosti, tačnije jedino eksperimentalna metoda ispunjava uslove neophodne za dokazivanje uzročnosti. Filozof znanosti, John Stuart Mill je formulisao tri uslova za dokazivanje uzročnosti koji se danas koriste kao standard. Tri uslova su:

1. Vremenski slijed (uzrok vremenski mora prethoditi posljedici)
2. Povezanost (uzrok i posljedica moraju biti povezani)
3. Ostala moguća objašnjenja povezanosti uzroka i posljedice moraju biti eliminirana

Eliminiranje ostalih mogućih objašnjenja odnosa uzroka i posljedice je jedino moguće primjenjujući eksperimentalni nacrt tj. eksperimentalnu kontrolu. Neki od postupaka eksperimentalne kontrole su uvođenje kontrolne grupe, uravnotežavanje, kontrola uslova i sl.

Varijable u eksperimentu mogu biti kvalitativne ili kvantitativne. Kvalitativne varijable označavaju osobinu koja se ne može brojčano izraziti (npr. boja očiju) dok kvantitativne označavaju pojave, tj. osobine koje se mogu brojčano izraziti. Kvantitativne varijable se dalje dijele na diskretne i kontinuirane. Diskretne varijable mogu imati samo određene vrijednosti na kontinuumu (cijelobrojne vrijednosti) dok kvantitativne mogu imati bilo koju vrijednost.

Eksperimentalni postupak podrazumijeva manipulaciju jednom varijablom, kako bi se mjerili efekti na drugu varijablu. Varijabla kojom se manipulira naziva se nezavisna varijabla, dok se varijabla na koju se mjere efekti naziva zavisna. Kada govorimo o eksperimentima u kojima se testira upotrebljivost ili neki njeni aspekti, nezavisna varijabla je aplikacija koja se

evaluirati. Recimo da nas zanima kako naš proizvod kotira kod korisnika u odnosu na konkurenciju. Nezavisna varijabla bi imala dva nivoa, tj. naš i konkurentski proizvod, dok je zavisna varijabla npr. korisničko zadovoljstvo ili stopa uspješnog završetka itd. Cilj eksperimenta je ispitati odnos zavisne i nezavisne varijable, što nužno podrazumijeva i izoliranje tog odnosa od djelovanja drugih faktora tj varijabli. Stepem u kojem možemo promjene na zavisnoj varijabli pripisati djelovanju nezavisne varijable naziva se unutarnjom valjanosti. Kako bi se postigla visoka unutarnja valjanost, neophodno je primijeniti postupke eksperimentalne kontrole. Generalno, nije moguće postići apsolutnu kontrolu, tj. apsolutno izolirati odnos nezavisne i zavisne varijable, ali je potrebno preduzeti sve mjere da bi se logički moguća alternativna objašnjenja odnosa eliminisala. Neke od tih mjera podrazumijevaju izjednačavanje ispitanika u grupama s obzirom na relevantne karakteristike, izjednačavanje uslova u grupama i sl.

4.1. Eksperimentalni nacrti sa nezavisnim grupama

Eksperimentalni nacrti sa nezavisnim grupama su najčešće korišteni eksperimentalni nacrti. U ovim istraživanjima uporedimo minimalno dvije skupine ispitanika koje su stavljene u različite uslove (broj grupa je određen nivoima nezavisne varijable). Uslove u svakoj od grupa kontrolira eksperimentator. Kako bi ova istraživanja bila valjana, potrebno je da su skupine izjednačene u uslovima i strukturi ispitanika, te da je jedina razlika eksperimentalna manipulacija nezavisnom varijablom. Kada se postigne izjednačenost grupa u svim relevantnim karakteristikama, onda sve eventualne izmjerene razlike na zavisnoj varijabli se mogu pripisati razlikama u nivoima nezavisne varijable koja djeluje.

Kod nacrti sa nezavisnim grupama razlikujemo kontrolnu i eksperimentalnu grupu. U kontrolnoj grupi imamo ispitanike koji nisu izloženi djelovanju nezavisne varijable. Prema Milas (2005) razlikujemo: netretiranu kontrolnu grupu, kontrolnu grupu standardnog tretmana, kontrolnu grupu nespecifičnog tretmana te kontrolnu grupu sparenih ispitanika. Mi ćemo se ovdje osvrnuti na netretiranu kontrolnu grupu i kontrolnu grupu standardnog tretmana zbog toga što imaju najširu primjenu u testiranju upotrebljivosti.

Kada govorimo o netretiranoj kontrolnoj grupi, jedina razlika između kontrolne i eksperimentalne grupe je prisustvo „tretmana“ (ovdje pod pojmom tretman podrazumijevamo prisustvo neke intervencije, izmjene ili u slučaju testiranja upotrebljivosti korištenje nekog IT alata ili ne). Recimo, ukoliko nas zanima upotrebljivost aplikacije za praćenje gradskog prevoza u Sarajevu, eksperimentalna skupina bi koristila aplikaciju dok bi kontrolna postupala bez ikakvih informacija o tome kada naredni autobus dolazi.

Kontrolna grupa standardnog tretmana podrazumijeva da ispitanike u kontrolnoj skupini izložimo uobičajenom postupku. Ovaj nacrt se najviše koristi prilikom testiranja upotrebljivosti jer u sustini podrazumijeva da kontrolna skupina koristi trenutni alat za neki zadatak. Npr. ukoliko testiramo edukacijski softver, kontrolna skupina bi istom sadržaju bila podučavana na klasični način u učionici, ili bi koristila trenutno najrasprostranjeniji IT alat za savladavanje tog sadržaja.

4.1.1. Odabir ispitanika za eksperiment i raspored po grupama

Kao što je već napomenuto pitanje odabira ispitanika je jedno od ključnih u planiranju istraživanja. Kada govorimo o eksperimentalnim nacrtima, reprezentativnost uzorka nije jedini faktor o kojem trebamo misliti. U cilju donošenja valjanih zaključaka potrebno je preduzeti različite korake kako bi se osigurala izjednačenost eksperimentalne i kontrolne grupe. Ujednačenost ispitanika po grupama, s obzirom na sve relevantne karakteristike je ključna za mogućnost pripisivanja eventualnim izmjerenih razlika djelovanju nezavisne varijable. Uzmimo za primjer da želimo testirati upotrebljivost novog knjigovodstvenog programa koji nudimo. Eksperimentalna grupa bi koristila naš novi program a kontrolna bi koristila trenutno najrašireniji program koji se koristi u te svrhe. Ukoliko skupine ne izjednačimo s obzirom na varijable kao što su: godine života, godine radnog iskustva, nivo IT vještina, eventualne razlike u rezultatima nećemo moći sa sigurnošću pripisati razlici u korištenom IT alatu. Drugim riječima, naše istraživanje će imati nisku unutarnju valjanost. Kao što je prethodno objašnjeno, vanjska valjanost se odnosi na mogućnost generalizacije rezultata istraživanja na kompletnu populaciju i relevantne situacije, dok se unutarnja valjanost odnosi na stepen sigurnosti u zaključivanju da su izmjerene razlike na zavisnoj varijabli rezultat manipulacije nezavisnom varijablom. Ukoliko imamo neizjednačene kontrolnu i eksperimentalnu grupu, unutarnja valjanost je ugrožena. U prethodno opisanom primjeru, razlike u uspješnosti korištenja našeg alata se mogu pripisati razlikama u nivou IT vještina ispitanika po grupama. Moguće je da u eksperimentalnu skupinu izaberemo dominantno mlađe ispitanike koji se bolje snalaze sa IT tehnologijom, što će ih učiniti uspješnijim u izvršavanju zadataka od ispitanika u kontrolnoj skupini. U tom slučaju, izmjerene razlike u uspješnosti ne možemo isključivo pripisati superiornosti našeg programa, nego nekim drugim faktorima. Dakle, u ovom slučaju, unutarnja valjanost našeg istraživanja je niska.

Postupci kojima osiguravamo ujednačenost ispitanika po grupama su (Milas 2005):

1. slučajni odabir
2. raspoređivanje po slučaju (randomizacija)
3. uparivanje.

Slučajni odabir je postupak koji podrazumijeva da svaki pripadnik populacije od interesa ima jednaku vjerovatnoću pojavljivanja u uzorku u svakoj od grupa. U praksi je jako teško osigurati da svi pripadnici populacije od interesa imaju jednaku vjerovatnoću da uđu u uzorak. U našem primjeru populacija od interesa bi bili svi pojedinci koji koriste knjigovodstvene programe u našoj zemlji (ili šire). Sa praktične strane jako je teško osigurati jednaku vjerovatnoću da svi budu dio našeg istraživanja. U slučaju da možemo osigurati jednaku vjerovatnoću za sve pripadnike populacije, ispitanike za svoj uzorak biramo metodom slučajnog odabira (recimo izvlačenje papirića i sl). Na ovaj način ne osiguravamo samo da svi pripadnici imaju jednaku vjerovatnoću da uđu u uzorak, nego i da imaju jednaku vjerovatnoću da budu svrstani i u eksperimentalnu i kontrolnu skupinu. Ipak, iako je u teorijskom smislu ovaj način uzorkovanja najidealniji, gotovo nikad se ne koristi. Izuzetak su studije u kojima je populacija od interesa izuzetno mala i dostupna.

Kada govorimo o postupku raspoređivanja po slučaju, svaki ispitanik koji je uzet u uzorak ima jednaku vjerovatnoću da bude smješten u eksperimentalnu ili kontrolnu grupu. Ovaj postupak osigurava statističko izjednačavanje varijabli individualnih razlika po grupama. Naime, slučajnim rasporedom ispitanika po grupi istraživač sebi pravi svojevrsnu policu osiguranja koja mu osigurava zadovoljavajuću unutarnju eksperimentalnu valjanost. Naravno da ne možemo očekivati da grupe budu savršeno izjednačene (npr. nekada kada primjenjujemo tehniku slučajnog dodjeljivanja brojeva jedan broj se pojavi više puta što dovede do neujednačenosti broja ispitanika u grupama).

Uparivanjem se ispitanici u grupe razvrstavaju prema istovjetnosti relevantnih osobina. Ovim postupkom se izbjegava neujednačenost grupa do koje može doći i prilikom slučajnog raspoređivanja u grupe. Do pojave neujednačenosti dolazi najčešće u slučajevima kada imamo mali uzorak. Jedna od opasnosti istraživanja na malom uzorku je velik varijabilitet individualnih razlika koje je onda teško izjednačiti slučajnim odabirom. Uparivanje podrazumijeva svrstavanje ispitanika sa nekom određenom osobinom, izraženom u poznatom intenzitetu u jednu grupu, te smještanje ispitanika sa istom osobinom izraženom u što sličnijem intenzitetu u drugu grupu. Ovaj postupak se često koristi kod kliničkih istraživanja. Ako se ispituje uspjeh novog operativnog zahvata, za svakog pacijenta koji prođe zahvat odabere se pacijent iste anamneze - starosti, spola, prethodnih bolesti itd. koji je podvrgnut uobičajenom operativnom zahvatu. Posebno interesantan primjer uparivanja je u studiji koja analizira uticaj osobina vozača (nezavisnih varijabli) na uzrokovanje sudara (zavisna varijabla) gdje su u eksperimentalnoj grupi vozači koji su skrivili sudar, a u kontrolnoj vozači drugog auta (Perneger i Smith, 1991). Na ovaj način je eliminisan uticaj drugih povezanosti kao što su vremenski uslovi, vidljivost, itd.

4.2. Vrste grupnih eksperimentalnih nacrt

4.2.1. Nacrt sa kontrolnom grupom i opažanjem prije i poslije intervencije

Općenito, ovdje se uglavnom radi o nacrtu u kojem imamo eksperimentalnu i kontrolnu skupinu, gdje kod kontrolne skupine nije prisutna intervencija (npr. rade posao na uobičajen način bez IT alata), dok je kod kontrolne prisutna intervencija čiji nas utjecaj zanima. Pri tome, zavisnu varijablu mjerimo i prije i poslije intervencije u obje skupine. U ovom slučaju možemo zaključivati o promjeni u eksperimentalnoj skupini prije i poslije intervencije, ali i s obzirom na postojanje ili ne postojanje razlike između eksperimentalne i kontrolne grupe. Ovakvi nacrti su nam korisni kada npr. želimo evaluirati uspješnost edukacijskog programa u sticanju znanja ili vještina.

Nacrt sa kontrolnom grupom i opažanjem prije i poslije intervencije ima široku primjenu jer je koristeći ovakav nacrt sa značajnim stepenom sigurnosti moguće zaključivati o veličini promjene do koje je dovela intervencija.

4.2.1.1. Statistički postupci koji se koriste prilikom analize rezultata nacrt sa kontrolnom i eksperimentalnom grupom i opažanjem prije i poslije intervencije

Česta greška koju istraživači prave prilikom statističke obrade rezultata ovih nacrt je da računaju t-test za prije i poslije mjerenje za eksperimentalnu i isto tako t-test za prije i

poslije mjerenje u kontrolnoj skupini. Do pogrešnog zaključka dolazi kada na osnovu statistički značajnog t-testa u eksperimentalnoj i neznačajnog testa u kontrolnoj skupini zaključimo da intervencija djeluje. U ovom slučaju bi ispravniji postupak bio poređenje razlika između grupa u prvom pa onda i u drugom mjerenju. Često se zna desiti da iako postoji razlika i pred i post mjerenju po grupama, nemamo statistički značajnu razliku između grupa ni na pred ni post mjerenju (Milas, 2005). Napominjemo da u slučaju poređenja grupa u prvom i drugom mjerenju, ne očekujemo postojanje razlike između grupa u prvom mjerenju (dakle mjerenju prije intervencije). Ukoliko nakon izmjerene statističke neznačajnosti u prvom, izmjerimo statistički značajnu razliku između eksperimentalne i kontrolne skupine, onda možemo govoriti o dokazanim efektima intervencije.

4.2.2. Nacrt sa kontrolnom grupom i mjerenjem samo poslije intervencije

Kod nacrt sa kontrolnom grupom i mjerenjem samo poslije tretmana, postupak odabira ispitanika u grupe i eksperimentalne kontrole je isti kao i kod prethodno opisanih nacrt. Razlika je u nedostatku mjerenja prije uvođenja intervencije. Ovaj nedostatak na prvu čini ovakve nacрте slabijim od gore opisanih nacrt, ali generalno, to ne mora biti slučaj. Naime, ukoliko sprovedemo sve postupke eksperimentalne kontrole za osiguravanje unutarnje valjanosti, pred mjerenje nije neophodno. Kako bismo osigurali unutarnju valjanost, potrebno je izjednačiti eksperimentalnu i kontrolnu grupu. Kod ovih nacrt mi nemamo informaciju o tome da li smo to i učinili tj. nemamo potvrdu koju imamo kada u pred mjerenju imamo statistički neznačajan t-test između grupa. Ipak, ukoliko smo sve uradili kako treba, možemo se osloniti na to da su grupe izjednačene i bez tog podatka. Ovdje je potrebno ponovo napomenuti da u slučajevima malih uzoraka nije realno očekivati potpunu izjednačenost grupa bez obzira na to koji je postupak izjednačavanja korišten.

Izostanak mjerenja prije intervencije pomaže kontroliranju efekta senzibiliziranja ispitanika. U slučaju kada primjenjujemo postupak pred mjerenja ispitanici dobiju svojevrsnu informaciju o tome šta mjerimo i naprave neku pretpostavku o tome šta se očekuje od njih. Ta pretpostavka može utjecati na njihovo ponašanje u toku eksperimenta, kao i na rezultate post mjerenja na različite načine. Ispitanici mogu odlučiti „pomoći“ istraživaču na način da će se ponašati onako kako misle da je „poželjno“ ili mogu prilikom drugog mjerenja pokušati biti konzistentni sa onim što su izjavili u prvom mjerenju.

Iako se to ne bi reklo na prvu, nacrt sa mjerenjem poslije tretmana je u određenoj prednosti u odnosu na nacrt sa mjerenjem prije i poslije tretmana (Milas 2005), osim u slučaju kada imamo razloga da sumnjamo u izjednačenost grupa. Nacrti sa mjerenjem poslije tretmana nisu odgovarajući u situacijama kada nam je potrebna informacija o veličini promjene do koje je došlo eksperimentalnom intervencijom (jer nemamo ulazne podatke).

4.3. Eksperimentalni nacrti sa ponovljenim mjerenjima

Za razliku od nacrt sa više grupa, gdje su ispitanici stavljeni u samo jednu situaciju (s obzirom na djelovanje nezavisne varijable), kod nacrt unutar ispitanika ista skupina ispitanika je izložena svim eksperimentalnim uslovima. Kako se onda mjerenje ponavlja u više vremenskih tačaka, ovi nacrti su nazvani nacrti sa ponovljenim mjerenjima. Tom

prilikom se posmatra ponašanje i pokušavaju detektirati eventualne promjene ponašanja ispitanika kroz različite uslove.

Kod nacrti sa ponovljenim mjerenjima ne moramo se brinuti o tome da li su skupine izjednačene, te nemamo tu opasnost za unutarnju valjanost. Dakle, kod nacrti sa ponovljenim mjerenjima ispitanici su sami sebi kontrola („kontrolna grupa“) što znači da se varijansa osobnih razlika u potpunosti kontrolira. Ukoliko ustanovimo promjenu na ponovljenom mjerenju nakon promjene nezavisne varijable ne brinemo o tome da li je razlika nastala kao rezultat razlika u ponašanju grupa. Zahvaljujući tome, eventualna alternativna objašnjenja promjena na zavisnoj varijabli su manje vjerovatna, te sa većim stepenom sigurnosti ih možemo pripisati djelovanju nezavisne varijable. Očekujemo da će vrijednosti na mjerama zavisne varijable varirati u skladu sa promjenama koje eksperimentator pravi na nezavisnoj varijabli.

Iz gore navedenog proizilazi da nacrti sa ponovljenim mjerenjima imaju veću statističku snagu od nacrti sa više grupa. Kod nacrti sa više grupa potrebno je dokazati promjenu kod različitih ispitanika u različitim uslovima i pripisati je djelovanju nezavisne varijable. Nacrti sa ponovljenim mjerenjima podrazumijevaju da dokažemo promjenu kod istih ispitanika u različitim uslovima, te da je pripišemo djelovanju nezavisne varijable.

Nacrti sa ponovljenim mjerenjima su upravo zbog veće statističke snage pogodni i za istraživanja kod kojih očekujemo male promjene u mjerama zavisne varijable tj. male efekte. Ekonomičnost je također na strani nacrti sa ponovljenim mjerenjima. Imamo manji broj ispitanika koje pratimo, mjerimo i sl. što dovodi do manjih materijalnih troškova.

Na osnovu gore opisanih prednosti nacrti sa ponovljenim mjerenjima možemo zaključiti o superiornosti ovih nacrti u odnosu na nacrti sa više grupa. Ipak, i ovi nacrti imaju nekih nedostataka. Npr. ovakvi nacrti su neizvodljivi u situacijama kada jedna eksperimentalna situacija trajno mijenja ispitanike kao što je slučaj u istraživanjima usvajanja vještina. Recimo da želimo testirati našu aplikaciju za učenje geometrije poredeći uspješnost poduke našom aplikacijom u odnosu na uspješnost koju ima konkurentska aplikacija. Ukoliko naše ispitanike naučimo koristeći jednu od aplikacija kako opisati kružnicu oko trougla, znanje i iskustvo kroz koje su ispitanici prošli ih nepovratno mijenja. U tom slučaju, dati istim ispitanicima isti zadatak na drugoj aplikaciji nije opravdano. Problem sa davanjem drugog zadatka je mogućnost da su zadaci tj. vještine koje treba savladati različite težine što nam narušava unutarnju valjanost. Ovdje bi se također postavilo i pitanje reda kojim su aplikacije date ispitanicima, zbog različitih mogućnosti utjecaja upravo vremenskog slijeda na rezultate (zamor ispitanika, poznatost i uvježbanost nakon prve situacije i sl.)

Pored toga, iako nacrti sa ponovljenim mjerenjima karakterizira manji broj ispitanika nego nacrti sa paralelnim grupama, obično se ispitanicima u ovim istraživanjima bavimo vremenski duže i intenzivnije. U tom slučaju često dolazi do osipanja ispitanika zbog zamora ili monotonije.

Kod nacrti sa ponovljenim mjerenjima nemamo prijetnje unutarljivoj valjanosti vezane za eventualne razlike u grupama, ipak postoje druge opasnosti o kojima moramo voditi računa te primjenjivati postupke eksperimentalne kontrole. Najveća prijetnja unutarljivoj valjanosti

ovih nacrtu se krije u redosljedu zadavanja eksperimentalnih situacija. Testiranje samo po sebi može utjecati na ispitanike na način da promijeni njihovo ponašanje ili rezultate na nekim mjerenjima. Sličan efekat imaju i promjene do kojih može kod ispitanika zahvaljujući protoku vremena između mjerenja (ovdje se misli na promjene koje nemaju direktne veze sa manipuliranjem nezavisne varijable). Ispitanici se s protokom vremena navikavaju na samu eksperimentalnu situaciju, postaju opušteniji (što može na različit način utjecati na njihov uradak), te mogu početi osjećati umor, dosadu i zasićenost.

Svi ovi faktori mogu utjecati na rezultate u smislu da imamo promjenu između mjerenja koju ne možemo pripisati sa sigurnošću djelovanju nezavisne varijable. U ovom slučaju onda govorimo o efektu redosljeda. Način za prevazilaženja ovog problema je u uravnoteženju uslova u svim eksperimentalnim situacijama. Postoje različiti načini na koje se postiže uravnotežavanje, ali se svi mogu podvesti pod zadavanje različitog rasporeda eksperimentalnih situacija (dakle ne situacija A pa situacija B, nego raspored, tamo gdje je moguće A, B, B, A ili A, B, A, B).

Kada govorimo o eksperimentima testiranja upotrebljivosti često je korisniji postupak kontrole permutacija eksperimentalnih uslova. Npr. ukoliko ispituje samo dva uslova (recimo dvije aplikacije) pola ispitanika bi imala jedan A,B raspored a druga polovica obrnut B,A. Istu kombinatoriku je potrebno napraviti ukoliko imamo tri i više eksperimentalnih situacija. Problem koji nastaje kada imamo više situacija se ogleda u tome da raste broj potencijalnih kombinacija koje treba uravnotežiti što značajno povećava i broj ispitanika koji su potrebni za istraživanje.

4.4. Složeni eksperimentalni nacrti

Kod eksperimentalnih nacrtu sa eksperimentalnom i kontrolnom skupinom kao i kod nacrtu sa ponovljenim mjerenjima analiziramo utjecaj jedne nezavisne varijable na zavisnu. Složeni, ili kako se još zovu faktorijalni nacrti ispituju utjecaj dvije ili više nezavisnih varijabli. Recimo da nas zanima uspješnost u razvoju vještina uz pomoć naše nove edukativne igrice za djecu. U jednom istraživanju možemo ispitivati da li se djevojčice ili dječaci bolje snalaze tj, imaju veći uspjeh u korištenju igrice. Drugo istraživanje možemo postaviti tako da ispituje utjecaj dječije dobi na uspješnost. Kod faktorijalnih nacrtu, te dvije nezavisne varijable: spol i dob ispituje istovremeno. Prednost ovog pristupa je u tome što dobijemo informacije o tome kako svaka od ove dvije varijable djeluje zasebno (glavni efekti), ali i kako djeluju u interakciji (efekt interakcije). Zahvaljujući informaciji o postojanju ili ne efekta interakcije, možemo doći do zaključka npr. da naša igrica daje najbolji efekt kod curica dobi od 10 do 12 godina, a najslabiji kod dječaka iste dobi. Daljnim ispitivanjem i analizom možemo doći do odgovora zašto je to tako te na koji način unaprijediti proizvod i napraviti strategiju plasmana na tržište.

Kada govorimo o analizi glavnih efekata, postupci koje primjenjujemo su slični onima iz jednostavnih eksperimenata. Cilj je odgovoriti na pitanje da li su promjene na zavisnoj varijabli rezultat djelovanja nezavisne varijable. U nacrtima koji uključuju više nezavisnih varijabli mjerimo glavni efekat svake od varijabli posebno. Dakle, koliko imamo nezavisnih varijabli toliko imamo i glavnih efekata.

Efekat interakcije nam daje informaciju o postojanju međudjelovanja, zajedničkog učinka dvije i više varijabli. O interakciji možemo zaključiti ukoliko učinak jedne varijable ovisi o nivou druge (kao što u našem primjeru učinak varijable spol ovisi o nivou varijable dobi). Naravno da broj interakcija koje analiziramo ovisi o broju nezavisnih varijabli koje imamo u eksperimentu.

Faktorijalne nacрте označavamo uz pomoć nezavisnih varijabli i broja razina koje nezavisne varijable imaju. Tako je nacrt u kojem analiziramo utjecaj spola (muško i žensko, dakle dvije razine) i dobi (mlada, srednja i kasna životna dob) na zadovoljstvo novom računovodstvenom aplikacijom, označit ćemo kao 2x3 nacrt po broju nivoa dvije nezavisne varijable. Nacrt 2x3x4 znači da imamo tri nezavisne varijable sa po dva, tri i četiri nivoa.

4.5. Analiza varijanse za obradu rezultata eksperimentalnih nacрта

Statistički postupak koji koristimo prilikom analize rezultata faktorijalnih nacрта je složena analiza varijanse. Analiza varijanse (ANOVA) je postupak koji se generalno koristi prilikom analize postojanja statistički značajnih razlika između aritmetričkih sredina tri ili više grupa. Đapo, Đokić (2012) navode kako se prilikom postupka analize varijanse ukupni varijabilitet razlaže na manje dijelove i to onaj koji se pripisuje djelovanju nezavisne varijable, te varijabilitet reziduala (greške).

Najjednostavniji primjer analize varijanse je jednostavna ANOVA kod koje imamo jednu nezavisnu varijablu. Jednostavna ANOVA se koristi prilikom analize rezultata nacрта sa nezavisnim grupama (gdje imamo tri ili više grupa i samo post mjerenje) i nacрта sa ponovljenim mjerenjima (gdje imamo tri ili više mjerenja kod iste skupine ispitanika). U zavisnosti od nivoa nezavisne varijable razlikujemo dvo ili tro-smjernu analizu varijanse. Kod složene analize varijanse imamo više nezavisnih varijabli. Tamo gdje je jedna zavisna varijabla, tu analizu nazivamo univarijatnom, dok onu gdje je više zavisnih varijabli multivarijatnom.

Kao rezultat postupka analize varijanse dobijemo vrijednost koju nazivamo F omjerom. Kod jednostavne analize varijanse dobijemo jednu vrijednost F omjera, dok kod složene imamo vrijednosti F omjera za sve glavne i efekte interakcije. F omjer može biti statistički značajan ili ne. Statistički značajan F omjer znači da je efekat faktora za koji se F omjer računao značajan, te da izmjerene razlike u aritmetičkim sredinama nisu nastale djelovanjem greške. Kada govorimo o jednostavnoj analizi varijanse, F omjer nam daje informaciju o tome da negdje između grupa postoji značajna razlika, ali ne i između kojih. Može se desiti da npr. postoji značajna razlika između 1. i 2. grupe i 2. i 3. ali ne i između 1. i 3. grupe. Kako bismo ustanovili gdje nam se utvrđena statistička razlika javlja, potrebno je uraditi post-hoc postupak.

Kod složene analize varijanse imamo F omjer za svaki glavni efekat i sve efekte interakcije. Ovdje je moguće da se desi da su recimo glavni efekti faktora neznačajni, ali da je interakcija značajna. Taj nalaz znači da nezavisne varijable djeluju na zavisnu samo u slučaju kada djeluju zajedno. Alternativna mogućnost je da imamo značajne glavne efekte, a neznačajnu interakciju, što znači da jedna (ili više, u zavisnosti od toga koliko imamo

statistički značajnih F omjera) nezavisna varijabla djeluje na zavisnu, međutim da zajedno više ispitivanih nezavisnih varijabli ne djeluje zajedno značajno na zavisnu.

4.6. Odabir eksperimentalnog nacrt

Kao što se moglo vidjeti iz prethodnih poglavlja, svaki od opisanih eksperimentalnih nacrti ima i svoje prednosti i nedostatke. Dio istraživačevog posla je odabir adekvatnog nacrti, tj. nacrti koji će na optimalan način dati odgovore na postavljeno istraživačko pitanje.

Prva odluka koju istraživač donosi je da li će implementirati klasični, jednostavni nacrt ili složeni. Kao što je objašnjeno, složeni nacrt omogućava obuhvatniju analizu fenomena i nalazi su relevantniji za realne uslove. S druge strane, zahtjevniji je za provesti, te je podatke teže analizirati s obzirom na to da je potrebno koristiti složenije statističke metode. Jednostavni nacrti su jednostavni za primjenu i analizu. Druga odluka se odnosi na broj istraživačkih uslova kojima je ispitanik izložen. Ukoliko jednog ispitanika izložimo u više eksperimentalnih uslova, onda koristimo nacrt unutar ispitanika (ponovljena mjerenja). Tamo gdje ponovljena mjerenja nisu moguća iz praktičnih ili etičkih razloga, koristimo nacrt između ispitanika (grupne).

Nacrti sa ponovljenim mjerenjima imaju veću statističku snagu, zbog toga što je sam ispitanik svoja kontrolna skupina. Kod ovih nacrti je lakše identificirati i dokazati učinak nezavisne na zavisnu varijablu. Nacrti sa ponovljenim mjerenjima se ne mogu koristiti u slučajevima trajnih promjena na ispitanicima do kojih dovode neke nezavisne varijable.

Najosjetljiviji zadatak istraživača kod nacrti unutar ispitanika je raspoređivanje eksperimentalnih uslova kako bi se elimirao utjecaj redoslijeda na rezultate. Kod nacrti sa paralelnim grupama jedan od najosjetljivijih istraživačkih zadataka je raspoređivanje ispitanika u grupe kako bi se osigurala izjednačenost grupa.

4.6.1. Odabir statističkog postupka za obradu podataka

Nakon odluke o tome koji će nacrt biti primijenjen u istraživanju, pristupa se prikupljanju podataka. Dalje, prikupljene podatke je potrebno unijeti u neki od programa za statističku obradu i pristupiti njihovoj analizi. Odluka o tome koji će se statistički postupak koristiti zavisi od više faktora. Đapo, Đokić (2012) nude pregled faktora koje treba uzeti u obzir, te odluka koje se donose na osnovu postojanja tih faktora. Na osnovu ovih pregleda kreirani su algoritmi odlučivanja prikazani u tabelama 4.1. i 4.2. Dati su algoritmi odlučivanja za studije sa dvije i više grupa pošto je to preduslov za eksperiment. Algoritmi odlučivanja za jednu grupu su dati u sljedećem poglavlju koje se bave ne-eksperimentalnim studijama.

Kao što je vidljivo iz algoritama odlučivanja datih u tabelama 4.1. i 4.2., potrebno je u prvom redu razmotriti nivo mjerenja na kojem smo. Većina mjerenja koje uključuju skale su na ordinalnom ili intervalnom nivou mjerenja. Naredno što se pitamo je kakva nam je distribucija rezultata. S obzirom na to da li je distribucija rezultata normalna ili nije odlučujemo se za korištenje parametrijske ili neparametrijske statistike.

Tabela 4.1. Algoritam odlučivanja kod poređenja dvije grupe

Poređenje dvije grupe	Skala mjerenja			
	Nominalna	Ordinalna	Intervalna / Omjerna	
			Utvrđena ekstremna asimetrija? Odstupanje od normalne raspodjele?	
	NE		Pretpostavlja se homogenost varijanski?	
	Vezani rangovi		DA	NE
DA		DA	NE	
Složeni Hi-kvadrat test		Mann-Whitney ili Wilcoxon test	t-test (zajednička varijansa)	t-test (odvojene varijansa)

Tabela 4.2. Algoritam odlučivanja kod poređenja tri i više grupa

Poređenje tri i više grupa	Skala mjerenja		
	Nominalna	Ordinalna	Intervalna / Omjerna
			Odstupanje od normalne raspodjele?
	NE		NE
	Vezani rangovi		
DA		ANOVA	
Složeni Hi-kvadrat test		Kruskal-Wallis test	

4.7. Artefakt u eksperimentu

Milas (2005) navodi kako je svaki zaključak temeljen na pogrešnim pretpostavkama o tome šta je dovelo do promjene na nezavisnoj varijabli artefakt. Dakle, govorimo o eksperimentalnim rezultatima na koje su utjecali faktori koji nisu predmet istraživanja i/ili nisu eksperimentalno kontrolisani. U prethodnim poglavljima smo spomenuli dosta prijetnji eksperimentalnoj valjanosti o kojima istraživač mora brinuti prilikom planiranja istraživanja, međutim postoje još neke, naizgled manje prijetnje valjanosti koje je lako previdjeti a mogu uzrokovati nevaljane rezultate.

Jedna od tih prijetnji je odnos eksperimentatora i ispitanika. Ova prijetnja se odnosi na očekivanja o tome šta treba biti ishod eksperimenta. Nerijetko se dešava da je izmjereni nalaz

rezultat ispitanikovih očekivanja o tome kako se trebaju ponašati tj. šta se od njih očekuje, a ne stvarnog utjecaja nezavisne na zavisnu varijablu. Milas (2005) navodi kako je eksperimentalna situacija psihološki problem za sebe. Razlog za ovu tvrdnju je uvid da ispitanici u pravilu nastoje pogoditi svrhu eksperimenta (ukoliko im nije rečena) i ponašati se u skladu s onim što misle da su eksperimentatorova očekivanja. Eksperimentator, s druge strane, često nesvjesno ispitanicima daje i svojevrsne tragove o tome šta je to što očekuje od eksperimenta, što dodatno podržava artificijelne promjene u ponašanju koje ispitanici počnu iskazivati.

U bilo kojoj istraživačkoj situaciji ispitanici se ne ponašaju u potpunosti prirodno. Generalno, sama informacija o tome da smo na neki način promatrani, u bilo kom kontekstu mijenja naše ponašanje. U eksperimentalnoj situaciji ispitanici dodatno pokušavaju odgonetnuti šta je to cilj eksperimenta, te šta je u toj situaciji poželjno ponašanje. Ispitanici se u istraživanjima ponašaju poput altruista koji žele pomoći istraživaču. Pri tome pogrešno zaključuju da se ta pomoć ogleda u ponašanju na način na koji misle da eksperimentator očekuje i/ili je društveno poželjan. U nekim slučajevima ispitanici ne žele iskazati ponašanja koja doživljavaju društveno nepoželjnim zbog činjenice da su svjesni da ih se promatra, pa aktivnosti koje imaju u eksperimentu ne izvode na uobičajen način. Pored toga, neki ispitanici su toliko voljni da „udovolje“ eksperimentatoru da će se ponašati potpuno suprotno od onoga što bi inače ili dati izjave za koje doživljavaju da eksperimentator želi čuti bez obzira na to šta stvarno misle. U većini slučajeva je i teško razdvojiti koja od te dvije motivacije je prisutna ili dominantna jer su u većini slučajeva prisutne obje. Dakle, ispitanici i žele da se prikažu u dobrom svjetlu i da pomognu istraživaču udovoljavajući onome što oni percipiraju kao očekivanja.

S obzirom na sve gore navedeno, postavlja se kao jedan od ključnih imperativa zahtjev da ispitanici ne proniknu u zahtjeve situacije. Potrebno je naglasiti da je i ispravan i pogrešan zaključak do kojeg ispitanici dolaze jednako poguban za eksperimentalnu valjanost jer mijenja ponašanje ispitanika. Najbolji način da se izbjegnu promjene u ponašanju ispitanika vezane za očekivanja o zahtjevima situacije je da im se ne daju ikakvi nagovještaji o tome šta je svrha eksperimenta. Postoje različiti načini na koje se ovo može postići. Tamo gdje eksperimentalna situacija dozvoljava primjenjujemo tkz. slijepi eksperiment. U slijepom eksperimentu koristimo obmanu na način da ispitanike uvjerimo da su izloženi nekom tretmanu, kada ustvari nisu. Ovaj model se najčešće koristi u kliničkim ispitivanjima lijekova gdje se jednoj grupi ispitanika kaže da primaju određeni lijek a ustvari im se daje placebo.

Drugi način podrazumijeva obmanu ispitanika na način da im se pruži pogrešna informacija o tome šta je predmet istraživanja. U ovom slučaju ponašanje ispitanika će biti promijenjeno ali ne u aspektima koji su relevantni za naše istraživanje. Dodatna opcija je mjerenje nezavisne varijable vremenski odmaknuto od eksperimentalne varijable ili u okruženju koje nije eksperimentalno. Milas (2005) navodi kako je greške u eksperimentalnom zaključivanju koje su posljedica ponašanja eksperimentatora moguće podijeliti u dvije kategorije: interakcijske i neinterakcijske. Neinterakcijske se odnose na iskrivljivanje rezultata zbog pristrasnosti ili osobne jednačine eksperimentatora, a da se pri tome ne utječe na ispitanika.

Eksperimentatorove greške često idu u smjeru postavljenih pretpostavki te znače rezultate pomjerene prema eksperimentatorovim očekivanjima. Dakle radi se o pristrasnosti u samom promatranju. Moguće je također da bude prisutna i pristrasnost prilikom tumačenja rezultata. Pristrasnost u tumačenju rezultata je manje problematična od pristrasnosti u opažanju. Razlog za ovo je u tome što pristrasno tumačeni rezultati uvijek postanu predmetom kritike ostatka naučne zajednice, dok je pristrasnost u promatranju teže uočiti nakon što su rezultati prikupljeni i istumačeni.

Eksperimentator često nesvjesno usmjerava ponašanje ispitanika u smjeru vlastitih očekivanja, ili ispitanicima daje nesvjesno tragove glede vlastitih očekivanja. Ispitanici zbog gore opisanih tendencija da udovolje eksperimentatorovim očekivanjima, mijenjaju svoje ponašanje. Dakle, jedna od prijetnji unutarnjoj valjanosti koja se ogleda u tome što ispitanici mijenjaju svoje ponašanje u skladu sa onim što oni misle da da je očekivanje eksperimentatora, se može smanjiti ukoliko eksperimentator „sakrije“ vlastita očekivanja.

Neki od postupaka kontrole pristrasnosti koje je povezana sa eksperimentatorovim ponašanjem uključuju povećanje broja istraživača u implementaciji istraživanja, smanjenjem kontakata eksperimentatora i ispitanika, standardizacijom postupka i objektivizacijom mjerenja, primjenom dvostruko slijepog nacrtu (ni istraživač ni ispitanici ne znaju ko je izložen kojim eksperimentalnim uslovima) ili primjenom kontrolne grupe kojoj su očekivanja iskomunicirana.

Iz prethodno navedenog jasno je da sve što eksperimentalne nacрте čini značajnim za znanstveno relevantna istraživanja, upravo čini da ispitanici nisu svjesni i aktivni učesnici evaluacije upotrebljivosti, što je preduslov za evaluacije upotrebljivosti sa uključenjem korisnika. Zato je važno razumjeti da podjela na eksperimentalne i ne-eksperimentalne studije ne dijeli istraživanja na bolja i lošija, ili više i manje značajna, već su to dva različita pristupa, namijenjena različitim kontekstima. Posebno je važno sagledati specifičnosti i jednog i drugog pristupa da bi se u sljedećim koracima – a to je obrada i analiza podataka – izabrale odgovarajuće statističke metode i time obezbijedila valjanost i interpretabilnost rezultata.

4.8. Kvaziekperiment

U prethodnom poglavlju je objašnjeno zašto se eksperiment smatra najmoćnijom istraživačkom tehnikom, posebno za zaključivanje o uzročnosti. Porijeklo znanstvene snage eksperimenta leži u izolaciji odnosa između zavisne i nezavisne varijable, koja se postiže primjenom rigorozne eksperimentalne kontrole. Eksperimentalna kontrola ima za cilj eliminaciju svih varijabli koje potencijalno mogu utjecati na odnos koji nas zanima. Ova izolacija sa sobom povlači pored artificijelnosti uslova u kojima se eksperimentalno istraživanje provodi i značajan broj ograničenja u smislu vrste pitanja na koja možemo tražiti eksperimentalni odgovor. Takve situacije u npr. pitanje da li je neka društvena intervencija učinkovita, učinkovitost neke obrazovne intervencije i sl. Implementacija eksperimentalne metode u prirodnom okuženju nam daje odgovore koji su imaju puno bolju vanjsku valjanost (mogućnost generalizacije rezultata na druge, prirodnije uslove. Kako često primjena eksperimentalne kontrole nije u potpunosti moguća u prirodnim uslovima, istraživači koriste postupak koji se naziva kvaziekperiment. Kvaziekperiment ima dosta odlika pravog

eksperimenta, ali kako često nije moguće primijeniti sve neophodne postupke eksperimentalne kontrole, sigurnost u zaključivanju o kauzalnosti je nešto umanjena. Postupci eksperimentalne kontrole podrazumijevaju izjednačavanje grupa, strogo kontrolisanje uslova, randomizaciju slova i sl. Svaki postupak koji podrazumijeva odstupanje od ovih standarda, smatra se kvazieksperimentom.

U slučajevima kada je nezavisna varijabla neka osobina nad kojom istraživač nema kontrolu (npr. spol, inteligencija, religioznost) istraživanje je moguće provesti formirajući grupe s obzirom na različitu izraženost nezavisne varijable. Slična situacija je u slučajevima kada je nezavisna varijabla izloženost nekom faktoru ili iskustvu koje je potencijalno štetno za ispitanike (ratna dejstva, trauma i sl). Kako je u svim istraživanjima strogo zabranjeno izlagati ispitanike potencijalno štetnim utjecajima, ovakva istraživanja moguće je realizirati samo na ispitanicima koji su u životu iskusili djelovanje nezavisne varijable. U ovim situacijama jasno da ne može biti randomizacije, tj. slučajnog smještanja ispitanika u grupe, što je jedan od važnih postupaka eksperimentalne kontrole. Ovakva istraživanja bi bilo primjer kvazieksperimenta. S obzirom na to, mogućnost zaključivanja o uzročnosti je mala, te najčešće govorimo o povezanosti između zavisne i nezavisne varijable, te dodatnim statističkim i metodološkim postupcima pokušavamo izolirati odnose varijabli ili kontrolisati konfundirajuće varijable kako bismo mogli donositi kvalitetnije zaključke. Drugim riječima, pokušavamo povećati unutarnju valjanost eksperimenta. Neki od tih postupaka uključuju dodavanje većeg broj kontrolnih grupa koje nisu izložene djelovanju nezavisne varijable ili uvođenje većeg broja mjerenja prije i poslije tretmana.

5. Ne-eksperimentalne metode evaluacije upotrebljivosti

U prethodnom poglavlju su opisane prednosti eksperimenta (i srodnog kvazieksperimenta) te je eksperiment okarakteriziran kao istraživački postupak sa najvećom snagom. Ipak, većina istraživanja u društvenim znanostima nisu eksperimentalna.

Eksperiment nazivamo aktivnom metodom zbog činjenice da istraživač aktivno manipulira nezavisnom varijablom kako bi mjerio nastajuće promjene na zavisnoj. U ne-eksperimentalnim istraživanjima istraživač nema mogućnost manipuliranja nezavisnom varijablom, niti kontroliše ostale varijable koje mogu biti relevantne za odnos. Fenomeni koje izučavamo ne-eksperimentalnim istraživanjima se dešavaju u prirodnim uslovima na prirodan način. Iz ovog razloga ne-eksperimentalna istraživanja ne daju konkretne i decidne informacije o uzročno-posljedičnim odnosima između varijabli tj. ne daju nam odgovor o tome da li je nezavisna varijabla uzrokom promjena na zavisnoj varijabli.

Generalno, zadatak znanosti je da predviđa pojave koje izučava. Ako to konkretiziramo na npr. psihologiju, zadatak psihologije kao znanosti je da predviđa ljudsko ponašanje i doživljavanje. U termina istraživanja mjerenja upotrebljivosti, cilj je predvidjeti frekvenciju i uspješnost korištenja alata koji nudimo. Prvi korak ka tom cilju je razumijevanje odnosa između pojava. U eksperimentalnim istraživanjima možemo dijagnosticirati uzročno-posljedični odnos, dok u ne-eksperimentalnim možemo govoriti samo o postojanju veze između izučavanih varijabli (dakle ne možemo precizirati prirodu te povezanosti). Iz navedenog proizilazi da ne-eksperimentalnim metodama možemo dobiti prve informacije o prirodi odnosa koji nas zanima i na osnovu toga postaviti relevantne hipoteze o uzročnosti koje onda testiramo eksperimentalnim nacrtom. Inače, istraživanje nekog fenomena ima upravo ovakve korake. Prvo se nekom od ne-eksperimentalnih metoda prikuplja veća količina podatka kako bi se dobile prve relevantne informacije. Nakon obrade tako dobijenih informacija pravi se istraživački plan za eksperimentalno istraživanje, tamo gdje je eksperiment moguće provesti. Izuzetak su fenomeni za koje postoji već značajan korpus istraživanja i relevantne informacije koje nam omogućavaju postavljanje znanstvenih hipoteza koje onda odmah provjeravamo eksperimentom. Ovdje je potrebno napomenuti da postoje fenomeni koje je nemoguće eksperimentalno ispitivati, te u smo u tom slučaju prisiljeni koristiti isključivo ne-eksperimentalne metode.

Ne-eksperimentalne metode nam daju informaciju o povezanosti između varijabli. Nekada nam se čini, nakon što dobijemo informaciju o povezanosti da se zaključak o uzročnosti sam nameće te da nam je jasan. Trebamo se paziti da ne upadnemo u zamku ovakvog zaključivanja jer je mogućnost pogreške velika. Kako prilikom primjene ne-eksperimentalnih istraživanja nemamo mogućnost kontrole uslova u kojima ispitujemo odnose, niti uvijek možemo složiti vremenski slijed (uzrok mora vremenski prethoditi posljedi, a kod ne-eksperimentalnih istraživanja sve podatke skupljamo u jednoj

vremenskoj tački) ne možemo isključiti postojanje alternativnih objašnjenja ispitivanog odnosa i djelovanje drugih faktora na odnos koji izučavamo.

Naizgled paradoksalno, osnovna prednost eksperimentalne metode je i njen glavni nedostatak. Izoliranjem odnosa između varijabli, stvara se vještačko okruženje koje skoro nikada nemamo u prirodnim uslovima. U vještačkim uslovima, ponašanje ljudi se (kako je već objašnjeno u poglavlju o eksperimentu) uvijek u određenom procentu mijenja, a nekada i drastično. Iz toga proizilazi da mi možemo biti potpuno sigurno o uzročno-posljedični odnos koji smo ustanovili eksperimentalnom metodom, a da te rezultate ne možemo replicirati u prirodnim uslovima kada djeluje više faktora na odnos i ispitanike. Što bolje kontroliramo uslove u eksperimentu to smo sigurniji u zaključke do kojih dođemo i istovremeno imamo manje šanse da se takva situacija ponovi u realnoj situaciji (što znači veću unutarnju i manju vanjsku valjanost). Ukoliko bismo insistirali na vanjskoj valjanosti eksperimenta gubi se osnovni smisao same metode. Tako da ne trebamo gledati eksperimentalna istraživanja kao nešto što treba da bude replika stvarnosti u laboratoriji jer to nije moguće postići i kao rezultat možemo imati nisku i unutarnju i vanjsku valjanost. Cilj eksperimentalnog istraživanja je da nam ponudi relevantne informacije o odnosu između varijabli. Na osnovu tih nalaza, gradimo daljni istraživački plan koji može uključivati nove eksperimente ili ne-eksperimentalna istraživanja, sve u cilj dobijanja definitivnih zaključaka o odnosu koji nas zanima i prirodni faktora koji utječu na taj odnos.

Koliko god eksperiment bio moćna istraživačka metoda, nekada ga jednostavno nije moguće primijeniti zbog prirode fenomena koji nas zanima. Npr. većina istraživanja koja uključuju istraživanje osobina ličnosti ili povezanosti osobina ličnosti sa nekim ponašanjima ili promjenama kroz vrijeme se ne mogu eksperimentalno ispitivati nego se moraju ispitati korelacijskom metodom. Također, u slučajevima kada ne možemo udovoljiti zahtjevu slučajnog raspoređivanja ispitanika u grupe, ili kada ne možemo kontrolirati nezavisnu varijablu bolje je i ne pokušavati koristiti eksperiment jer postoji velika opasnost od donošenja pogrešnih zaključaka. Sve navedeno utiče na značaj ne-eksperimentalnih metoda u evaluaciji upotrebljivosti softvera, koje se u praksi često koriste.

Rad sa jednom grupom se koristi za estimaciju vrijednosti od interesa ili poređenje sa banchmarkom. Vrijednosti od interesa su vrijednosti metrika vezanih za korisničko iskustvo odnosno upotrebljivost. Već je pominjano vrijeme završetka zadatka, a to mogu biti i korisničke ocjene aplikacije. Kod estimacije parametara populacije korisnika koristimo se deskriptivnim istraživanjima koja imaju visoku snagu zaključivanja kada su dobro osmišljena. U tabeli 5.1. je prikazan algoritam odlučivanja za studije sa jednom grupom. Prikazani algoritam je baziran pregledu faktora koji vode proces odlučivanja dat u Đapo, Đokić (2012).

Za evaluacije upotrebljivosti neophodno je istaći da je u praksi najčešće korišteni instrument mjerenja korisničkog iskustva upitnik sa ordinalnim skalama mjerenja, i iako se u tom slučaju ne mogu primijeniti statistički testovi koje koristimo za druge skale, kao mjere korisničkog iskustva se koriste metrike bazirane na frekvencijama odgovora, odnosno procentima, i njihovim kombinacijama. Tako da se za ocjenu korisničkog iskustva koristi ukupan procenat odgovora slaganja (eng. *Percent Agree*), procenat odgovora maksimalnog

slaganja (eng. *Top Agree*), ili Net Promotor Score, koji se računa kao razlika ukupnog procenta odgovora slaganja i ukupnog procenta odgovora neslaganja.

Tabela 5.1. Algoritam odlučivanja kod studija sa jednom grupom

Jedna grupa	Skala mjerenja			
	Nominalna	Ordinalna	Intervalna / Omjerna	
			Dovoljno veliki uzorak i poznato σ	
			DA	NE
Jednostavni Hi-kvadrat test	Top Agree*, Percent Agree*, Net Promotor Score*	z-test	t-test	

Posebno pitanje, prilikom razmatranja eksperimenta kao istraživačkog postupka je pitanje etičnosti istraživanja. Nekada je eksperimentalno istraživanje izvodljivo praktično, ali nije etički opravdano. Prilikom razmatranja izvodljivosti nekog istraživačkog postupka važno je da razmotrimo etički imperativ da se ispitanicima ne smije načiniti šteta zbog provođenja istraživanja. Neki od didaktičkih primjera praktično izvodljivih ali neetičnih istraživanja uključuju ispitivanje utjecaja radijacije na kognitivne sposobnosti, ili općenito izlaganje ispitanika stresorima ili traumatskim događajima u svrhu izučavanja različitih efekata istih. Nekada je ova istraživanja moguće realizirati koristeći kvaziekperimentalni nacrt (npr. stanovništvo na području Bosne i Hercegovine je bilo predmetom mnogobrojnih istraživanja nakon ratnih dejtava, posebno iz domena kliničke psihologije).

5.1. Posmatranje/opažanje

Posmatranje je aktivnost koju mi svi obavljamo svakodnevno. Iz tog razloga se na prvu možemo zapitati kako je to znanstvena metoda. Potrebno je stoga, odmah naglasiti da svako posmatranje nije znanstveno posmatranje. Dakle, aktivnost opažanja pojava i dešavanja oko nas koju mi obavljamo bez da posebno razmišljamo o tome šta radimo, nije znanstveno posmatranje.

Znanstveno posmatranje se od svakodnevnog posmatranja razlikuje suštinski u načinu na koji se provodi. Mi prilikom laičkog posmatranja, na osnovu dijela informacija koje procesiramo i izuzetno malog uzorka jedinki ili situacija koje posmatramo, dolazimo do zaključaka. U tom slučaju se izlažemo mogućnosti različitih grešaka u zaključivanju. Npr. moguće je da smo zamijetili samo jedan dio dešavanja kojem smo svjedočili, kao i da nam neke bitne informacije o akterima i samoj situaciji nisu dostupne. Također, u svakodnevnom promatranju mi donosimo zaključke na osnovu male veličine uzorka (pojedinaca i pojava). Ova činjenica uparena sa već spomenutim problemima dovodi do toga da često pravimo pogrešne zaključke.

Znanstveno promatranje, s druge strane u prvom redu karakterizira sistematičnost u pristupu. Znanstvenik koji posmatra to čini na velikom broju posmatranih pojava i

pojedinaca, pri čemu ono što opaža detaljno dokumentira tj zapisuje. Dakle, znanstveno posmatranje je sistematično i objektivno i to je ono što ga razlikuje od laičkog posmatranja. Rezultati se evidentiraju u skladu sa unaprijed definiranim planom i pravima kako bi se osiguralo prikupljanje svih relevantnih informacija.

Milas (2005) navodi kako istraživač obavlja izbor ponašanja koja opaža (najčešće prije, ali moguće je i tokom ili nakon opažanja). Ponašanja koja se rijetko javljaju u prirodnim uslovima moguće je izazvati ili javljanje tog ponašanja olakšati intervencijama na okolini. Kako bi se na osnovu velikog broja prikupljenih informacija mogli donositi valjani zaključci potrebno je uraditi kodiranje, tj razvrstavanje pojedinačnih ponašanja u unaprijed definirane kategorije.

Za razliku od nekih drugih metoda, metoda opažanja istraživaču omogućava neposredan uvid u ponašanje. Dakle, prilikom korištenja ove metode istraživač se ne oslanja na različite mjere samoiskaza ili posredne pokazatelje koji često mogu biti diskutabilni.

Pored toga, velika prednost metode opažanja je prirodnost uslova u kojima se odvija. Ostali istraživački postupci koji su nam dostupni podrazumijevaju (1) artifičijelnost situacije koja je najveća u eksperimentalnim istraživanjima, i/ili (2) oslanjanje na rezultate upitnika samoprocjene ili testova koji na određeni način ograničavaju tj pojednostavljaju stvarnost.

Ipak, slično kao i sa eksperimentalnim postupkom, najveća prednost metode opažanja je ustvari i najveći problem. Kao što je već objašnjeno, eksperimentalni postupak podrazumijeva izoliranje odnosa između varijabli od interesa na način da se eliminiraju svi drugi mogući utjecaji na taj odnos. Ovakva situacija je artifičijelna i gotovo nikada se ne dešava u prirodnim uslovima, zbog čega je izuzetno rijetko moguće replicirati rezultate eksperimenta van laboratorije. S druge strane, opažanje kao metodu karakterizira izuzetna prirodnost uslova. Problem nastaje kada želimo donijeti zaključke o uzročno – posljedičnim vezama ili općenito odnosima između konkretnih varijabli od mnoštva onih koje djeluju u svakom momentu. Iz ove razlike proizilaze i razlike u prirodi fenomena koji se izučavaju ovim metodama. Eksperimentom uglavnom ispituje fragmentirane dijelove fenomena, dok opažanjem ispituje ponašanja koja imaju širi smisao i daju nam uvid u sam proces.

Koristeći metodu opažanja istraživač traži odgovore na situacijske odrednice ponašanja tj faktore koji su povezani sa javljanjem fenomena koji nas zanima. Na osnovu tih nalaza, moguće je postaviti validne uzročno – posljedične hipoteze koje se onda mogu testirati eksperimentalnom metodom. Možemo zaključiti da je posmatranje metoda koja logički prethodi eksperimentu.

5.1.1. Vrste posmatranja

Metodu promatranja uglavnom razvrstavamo po tome u kojoj mjeri istraživač utječe na situaciju i uslove u kojima se dešava promatranje. S obzirom na taj faktor, razlikujemo promatranja sa i bez uključivanja. Milas (2005) navodi kako i razlikujemo promatranje cjelovitog ponašanja i promatranje izdvojene jedinice ponašanja.

Pored toga razlikujemo i vrste posmatranja s obzirom na stepen struktuiranosti, i to struktuiranosti okoline u kojoj se posmatranje obavlja i struktuiranosti podataka koji se u

samom postupku prikupljaju. Ako razmatramo struktuiranost okoline, onda opažanje može biti u u potpuno prirodnim uslovima ili laboratorijskim uslovima (i sve varijante između). S obzirom na strukturu sadržaja koji se prikuplja, posmatranjem možemo prikupiti sve informacije o situaciji koja se odigrala (ne razmatrajući relevantnost prilikom samog postupka prikupljanja) ili možemo imati unaprijed postavljenu strukturu u smislu unaprijed definiranih karakteristika i događaja koje opažamo i evidentiramo.

U određenom smislu, opažanje koje podrazumijeva visoko struktuiranje okoline, tj kontrolisanje okolinskih faktora koji djeluju na predmet opažanja, se primiče eksperimentu. Razlika je u stepenu kontrole i manipulaciji nezavisnom varijablom koja ne postoji kada govorimo o metodi opažanja.

5.1.1.1. Posmatranje bez uključivanja

Ovu vrstu promatranja karakterizira potpuno neuključivanje istraživača u fenomen koji posmatra. Prilikom primjene ovog postupka, istraživač je pasivni promatrač koji samo evidentira ono što se pred njim dešava. U teoriji, pojave koje se promatraju na ovakav način se odvijaju potpuno prirodno, onako kako bi se odvijale i bez prisustva istraživača. Prilikom analize podataka dobijenih na ovakav način nije moguće donositi zaključke o vezi između ponašanja i nekih situacijskih odrednica.

Promatranje koje se dešava na ovakav način, u smislu pasivnosti istraživača čije prisustvo nije detektirano od strane osobe/a koje se promatraju, je idealni slučaj koji je jako teško izvediv. U većini slučajeva osobe su svjesne istraživačevog prisustva, te on preduzima različite korake kako bi osigurao da njegovo prisustvo ne utječe na pojavu koju posmatra. U tu svrhu postoje različite tehnike koje se koriste, ali generalno je uspješnost istih vrlo upitna. Razlog za ovo leži u činjenici da se ljudi čak i nesvjesno ponašaju drugačije ukoliko znaju da su promatrani. U slučajevima kada se posmatra neka aktivnost čija se uspješnost mjeri, neki ispitanici osjećaju dodatni pritisak jer znaju da ih se posmatra kako izvode zadatke zbog čega njihov učinak opada. Ukoliko želimo da posmatramo ponašanje ljudi u nekim svakodnevnim situacijama onda se izlažemo riziku da se ispitanici ponašaju onako misle da treba, bilo da imaju pretpostavku o tome šta mi očekujemo za vidjeti ili o tome šta je društveno poželjno.

Pri tome, u situaciji kada želimo da sakrijemo prisustvo istraživača, izlažemo se ozbiljnoj etičkoj dilemi. U tom slučaju primjenjujemo manipulaciju ispitanika i nekada koristimo obmanu. Prilikom donošenja odluke o sprovođenju istraživanja na ovakav način moramo razmotriti da li postoji mogućnost nanošenja štete ispitanicima, te moramo imati adekvatan feedback za ispitanike koje smo promatrali po završetku postupka.

5.1.1.2. Posmatranje sa uključivanjem

Iako se posmatranje bez uključivanja čini optimalnim postupkom (upravo zbog prirodnosti uslova u kojima se odvija), ono često nije izvodivo. Naime, istraživača nekada zanimaju pojave koje se rijetko javljaju spontano ili nas zanima specifična manifestacija neke pojave. U oba slučaja, posmatranje u prirodnim uslovima je praktično neizvodljivo jer bi zahtjevalo ogroman broj istraživača koji bi duži vremenski period posmatrali veliki broj ljudi kako bi prisustvovali spontanom javljanju pojave koja nas zanima.

Iz ovih razloga istraživač koristi tehniku posmatranja sa uključivanjem, pri čemu onovi neke nove elemente u situaciju kako bi postigao:

- Javljanje ponašanja koje se rijetko javlja ili je javlja u situacijama u kojima ga je teško posmatrati
- Praćenje ponašanja u različitim situacijama (tj. kako bi na osnovu posmatranja istog ponašanja u različitim situacijama mogao postaviti hipoteze o situacijskim odrednicama ponašanja).

Milas (2005) dijeli posmatranje sa uključivanjem na: sudioničko posmatranje, strukturirano opažanje i eksperiment u prirodnim uslovima. Sudioničko opažanje podrazumijeva blisku interakciju istraživača sa sudionicima koji se posmatraju. U nekim slučajevima ispitanici koje posmatramo budu svjesni istraživačevog prisustva u svojstvu promatrača, a u nekim to nije slučaj. Problem se javlja u činjenici da svjesnost da smo promatrani mijenja ljudsko ponašanje. Zbog toga je često neophodno od ispitanika sakriti da su promatrani, kako bismo osigurali da se ponašaju na prirodan način.

Prilikom sudioničkog posmatranja istraživač doživljava slična iskustva kao i ispitanici što može dalje dovesti do poistovjećivanja sa njima. Ovdje se javlja potencijalni metodološki problem koji je vezan za smanjenu objektivnost istraživača kao i za selektivnu istraživačku percepciju zbog utjecaja vlastitog iskustva. Istraživač stoga mora biti svjestan ovih opasnosti i preduzeti pripremne radnje kako se to ne bi desilo.

Drugi problem vezan za sudioničko posmatranje vezan je za istraživačev utjecaj na samu situaciju koju promatra. Zbog činjenice da istraživač neupitno sudjeluje u pojavi koju promatra, moramo se zapitati da li bi se ta pojava odvijala na isti način i da istraživač nije nazočan. Zbog toga je dobro da istraživač što manje utječe na pojavu koju promatra i na taj način kontroliše mogućnost vlastite manipulacije (svjesne i nesvjesne) na rezultate. Dalje, upravo kako bi se otklonila ova opasnost, zaključci se ne donose na osnovu jednog posmatranja, nego na osnovu više posmatranja od strane više istraživača.

Za razliku od opažanja u prirodnim uslovima i sudioničkog opažanja gdje želimo utjecaj istraživača svesti na minimum, kod strukturiranog posmatranja unaprijed planiramo različite situacije koje želim posmatrati. Ilustrativni primjer ove vrste posmatranja su klinički uslovi gdje sa roditeljima dogovaramo set situacija u kojima ćemo ih posmatrati sa njihovom djecom kako bi se došlo do zaključaka o međusobnoj interakciji i dječijem ponašanju. Strukturirano opažanje se na površini čini sličnom eksperimentu, posebno u slučajevima kada se odvija u laboratorijskim uslovima. Ipak, značajna razlika između strukturiranog opažanja i eksperimenta je u manipulaciji nezavisnom varijablom koja se ne dešava u primjeni metode opažanja ali je neizostavan dio eksperimentalnog postupka.

5.1.1.3. Eksperiment u prirodnim uslovima

Eksperiment u prirodnim uslovima je postupak u kojem istraživač u prirodnim uslovima manipulira jednom ili sa više nezavisnih varijabli. Riječ je hibridnoj tehnici između eksperimenta i posmatranja. Prednost ovakvog postupka je veća prirodnost situacije u kojoj se efekti ispituju. S druge strane manja je mogućnost kontrole nad vanjskim uslovima. Ovoj

metodi pribjegavamo u situacijama kada je jako teško u laboratoriskim uslovima izazvati ponašanje koje nas zanima.

5.1.2. Vrste ponašanja koje je moguće oponašati

Izbor ponašanja koje će se promatrati je samo naizgled jednostavno i vrlo važno pitanje. Moderna tehnologija, posebno dostupnost i jednostavnost video zapisa je samo naizgled pojednostavila ovu situaciju. Naime, odabir ponašanja koje ćemo posmatrati je jednako važna odluka i jednako je u istraživačevim rukama kao i što je to bilo prije mogućnosti video zapisa. Pitanje ponašanja koje će biti posmatrano je ustvari pitanje konstruktne valjanosti istraživanja. Ponašanje ili aspekt ponašanja koji biramo posmatrati treba biti odraz konstrukta koji nas zanima. Očekujemo da je jedna skupina ponašanja povezana sa jednim aspektom teorijskog konstrukta koji nas zanima, dok drugu skupinu ponašanja povezujemo sa drugim aspektima. Pitanje tačnosti ovog zaključivanja je ustvari pitanje konstrukte valjanosti.

Prilikom odabira ponašanja koje ćemo posmatrati treba razmišljati i o pouzdanosti pokazatelja koje biramo. Svaki konstrukt je potrebno opisati većim brojem pokazatelja kako bismo bili sigurni u vlastite prosudbe. Milas (2005) navodi četiri skupine ponašanja koja dolaze u obzir za posmatranje i to: (1) neverbalno, (2) prostorno, (3) ekstralingvističko i (4) lingvističko.

Neverbalni pokreti su često bolji pokazatelj našeg psihičkog stanja od onoga što verbalno saopštavamo o svojim emocijama. U nekim situacijama naše neverbalno ponašanje je u nesuglasju sa onim što govorimo. Iz tog razloga, bilježenje i analiza našeg neverbalnog ponašanja je u najmanju ruku korisna nadopuna verbalnim informacijama koje ispitanici daju o svojim emocijama i psihičkim procesima. Problem sa kojim se surećemo prilikom analize neverbalnih znakova je u tome što se često radi o izuzetno suptilnim i teško uočljivim znakovima koji nam mogu promaći. Iz tog razloga, informacije koje dobijemo neverbalnom komunikacijom je dobro koristiti kao nadopunu informacijama koje prikupimo iz drugih izvora.

Međusobna udaljenost između ispitanika je također korisna dodatna informacija koju možemo prikupiti prilikom opažanja. Udaljenost do pola metra se smatra intimnim prostorom, od jedan do dva metra osobnim prostorom, do tri metra se obično obavljaju poslovni i društveni kontakti a preko tri metra se smatra javnim prostorom. Na osnovu ovih parametara možemo pretpostavljati nivo prisnosti između ispitanika.

Ekstralingvističko i lingvističko ponašanje se odnosi na jezičko, verbalno komuniciranje (lingvističko ponašanje) i na popratne aspekte jezičkog komuniciranja kao što je jačina i boja glasa, brzina govora i sl. (ekstralingvističko).

5.1.3. Analiza podataka prikupljenih posmatranjem

Posmatranjem u pravilu dobijemo veliki broj podataka koje je potrebno na neki način obraditi kako bi se mogli izvesti zaključci. Prvi korak u tom procesu je organiziranje informacija u manji broj osnovnih nalaza. Ovaj postupak se naziva kodiranjem. Kodiranje podrazumijeva klasifikaciju posmatrane pojave s obzirom na modalitete u kojima se javlja (npr. modalitet ponašanja djece u učionici). Postupak podrazumijeva identificiranje

ponavljajućih elemenata, definiranje kategorija na osnovu toga i onda razvrstavanje rezultata s obzirom na to kojoj kategoriji pripadaju. Nakon toga se pravi daljnja analiza s obzirom na frekvenciju javljanja ponašanja u pojedinoj kategoriji.

5.1.4. Objektivnost mjerenja

Objektivnost posmatranja se odnosi na pitanje da li različiti opažači vide isti događaj na isti način. Objektivnost se dakle izražava stepenom slaganja između više opažača. Tamo gdje nemamo objektivnost imamo problem sa uopćavanjem zaključaka do kojih smo došli.

Uzroke neobjektivnosti možemo tražiti u samim karakteristikama opažača, tj. u razlikama u iskustvu, utreniranosti, trenutnim stanjima (kao što je umor). Ovi problemi se prevazilaze treningom opažača ili davanjem povratne informacije o postojanju mimoilaženja. Drugi uzrok neobjektivnosti se ogleda u nepreciznom i nejasnom definiranju samog postupka opažanja. Potrebno je pojavu ili događaj koji se posmatra jasno i precizno operacionalizirati kako bi svi opažači imali jasnu sliku o tome šta posmatraju. Pored toga, potrebno je voditi računa i o tome da je operacionalizacija posmatranog ponašanja tačna, tj. da odgovara teorijskim postavkama i prethodnim saznanjima o posmatranom fenomenu. Ukoliko ovaj uslov ne ispunimo izlažemo se opasnosti da imamo objektivno, ali netačno mjerenje (stručno rečeno, nevaljano mjerenje jer u tom slučaju ustvari ne mjerimo ono što mislimo da mjerimo. Slaganje istraživača se izražava koeficijentima korelacije.

Utjecaj opažača na opaženo ponašanje je fenomen koji isto utiče na objektivnost mjerenja. U psihologiji je odavno poznato i dokazano da svjesnost o tome da smo posmatrani mijenja naše ponašanje. Ova činjenica ima značajne reperkusije na sam postupak provođenja istraživanja i potrebno je uzeti u obzir prilikom interpretacije rezultata.

5.2. Korelacijsko istraživanje

Uopćeno gledano, psihologijska istraživanje možemo podijeliti na dvije velike skupine istraživanja, eksperimentalna i korelacijska. Korelacijski pristup se bavi ispitivanjem povezanosti varijabli u prirodnom okruženju i osobnih razlika između pojedinaca. Eksperimentalni pristup individualne razlike nastoji odstraniti ili kontrolisati kako bi došli do saznanja o situacijskim odrednicama ponašanja. Dakle, korelacijskim pristupom se proučavaju razlike među ljudima, dok se eksperimentalnim pokušava doći do zaključaka o opštim zakonitostima.

Većina ne-eksperimentalnih istraživanja u psihologiji su ustvari korelacijska. Korelacijskim istraživanjem dobijamo informaciju o postojanju povezanosti između varijabli od interesa. U korelacijskom istraživanju ne pribjegavamo manipulaciji varijablama kako bismo mjerili efekte (kao i eksperimentalnim). Naprotiv, u korelacijskim studijama, u istoj vremenskoj tački, u prirodnim uslovima prikupljamo podatke o svim varijablama koje nas zanimaju. Informacije koje korelacijskim istraživanju dobijemo o povezanosti između varijabli nam omogućavaju i predviđanje rezultata. Milas (2005) navodi da se u psihologijskoj literaturi korelacijska istraživanja dijele na psihologijsko testiranje i anketiranje. Kako je u istraživanju upotrebljivosti u praksi najčešće zastupljeno anketiranje ovaj pristup

korelacijskim istraživanjima ćemo izdvojiti u posebno poglavlje da bi na odgovarajući način prodiskutovali sve važne odluke koje se trebaju donijeti kod provođenja ovog istraživanja.

5.2.1. Psihologijsko testiranje

Jedan od najčešćih načina prikupljanja podataka u korelacijskim istraživanjima podrazumijeva primjenu psihologijskih testova. Psihologijski testovi su posebno razvijeni, standardizirani instrumenti koji omogućavaju mjerenje najrazličitijih ljudskih osobina. Neke od osobina koje nas zanimaju i u istraživanjima upotrebljivosti su npr.: intelektualne sposobnosti, različiti stavovi, profesionalni interesi, osobine ličnosti i sl.

Prednost primjene psihologijskih testova u istraživanjima leži u činjenici da je mjerenje obavljeno na takav način objektivnije, valjanije i pouzadnije nego da primjenjujemo za potrebe istraživanja posebno dizajnirane skale. Naravno da je u praksi istraživanja upotrebljivosti često neophodno i dizajnirati skale kako bismo mjerili različite aspekte upotrebljivosti vezane za konkretnu situaciju koja nas zanima, ali generalno, primjena standardiziranih testova i skala je metoda izbora tamo gdje je moguće.

Kao što je objašnjeno i u poglavlju o mjernim instrumentima, karakteristike psihologijskih testova su: objektivnost, valjanost, pouzdanost i diskriminativnost. Prednost korištenja psihologijskih testova, kao i standardiziranih upitnika je upravo u tome što su ove karakteristike unaprijed izmjerene i poznate, što nam čini mjerenje i zaključke koje donosimo utemeljenijim i tačnijim.

5.3. Tumačenje nalaza korelacijskog istraživanja

Korelacijsko istraživanje je istraživački postupak koji nam omogućava da u relativno kratkom vremenskom roku, uz niske materijalne troškove skupimo veliku količinu podataka. Na osnovu podataka dobivenih korelacijskim istraživanjima možemo generirati puno hipoteza koje onda provjeravamo eksperimentalnim studijama. Primjenom više istraživačkih postupaka u cilju rasvjetljavanja nekog fenomena ili problema, dobijamo potpuniji uvid i sigurniji smo u zaključke koje donosimo zbog toga što se različite metode nadopunjuju.

Najveći problem s kojim se istraživači prilikom primjene korelacijskih istraživanja suočavaju je u tome što korelacijskim studijama možemo dokazati povezanost između varijabli, ali ne i definirati uzročno – posljedične veze.

Jedini istraživački postupak koji nam omogućava da govorimo u terminima uzroka i posljedica je eksperiment. Korelacijskim studijama mi dobijemo informaciju o postojanju povezanosti između varijable X i Y, ali ne i o tome da je npr. varijabla X uzrok varijabli Y. Na osnovu korelacijske studije ne možemo isključiti niti jedan od ova tri slučaja:

- 1) X je uzrok Y,
- 2) Y je uzrok X,
- 3) Z je uzrok X i Y .

Prve dvije opcije se tiču smjera povezanosti o čemu nam sama korelacija ne daje informaciju. Ukoliko npr. utvrdimo postojanje povezanosti između igranja kompjuterskih

igrice koje imaju nasilan karakter i nasilnog ponašanja kod djece, ne možemo na osnovu korelacijskog istraživanja reći da li je agresivni sadržaj u igricama uzrokom agresivnog ponašanja, ali djeca agresivnog ponašanja biraju agresivne sadržaje. Nekada se smjer povezanosti čini logičan i u tim situacijama moramo biti izuzetno oprezni prilikom eliminiranja jedne od opcija jer iako nam se nekada čini da je situacija jasna, postoji i mogućnost obrnutog smjera.

Treći slučaj se odnosi na mogućnost isključivanja utjecaja treće varijable na odnos varijabli sa kojima postoji povezanost. U slučaju postojanja utjecaja treće varijable moguće objašnjenje povezanosti tj. izmjerene korelacije između varijabli X i Y je upravo u tome, da nam mjerenje pokazuje da postoji povezanost zbog djelovanja trećeg, zajedničkog faktora. Uzmimo za primjer postojanje povezanosti između varijable dob i varijable lična primanja. Logičnim se čini zaključak o smjeru povezanosti u smislu da veća životna dob doprinosi većim ličnim primanjima. Međutim, korelacijskim istraživanjem nismo eliminirali treću varijablu koja moderira odnos ove dvije, a to je iskustvo. Dakle, korelacijskim istraživanjima, posebno jednim korelacijskim istraživanjem nije moguće govoriti o uzročnosti. Postoje načini da se korelacijska istraživanja osnaže kako bi se mogle praviti razumne pretpostavke o uzročnosti ali u tom slučaju govorimo o naprednim statističkim i metodološkim postupcima.

5.4. Anketno istraživanje

Anketno istraživanje je poseban oblik istraživanja u kojem se ispitanike direktno pita o njihovim mišljenjima, stavovima, ponašanju, vrijednostima i sl. Ove podatke prikupljamo posebno dizajnirajući skup pitanja u te svrhe. U anketnom istraživanju ne posmatramo direktno ponašanje naših ispitanika, nego ih pitamo o tome kako se ponašaju (tipično, u određenim situacijama i sl). Anketna istraživanja se mogu dizajnirati u svrhu opisa, istraživanja ili objašnjavanja pojava. Opisna istraživanja imaju za cilj prikupljanje informacija o rasprostranjenosti neke pojave. Ukoliko pored dijagnostike rasprostranjenosti neke pojave, istraživanjem želimo ući u suštinu i detaljno je opisati, onda govorimo o anketama koje ima za cilj istraživanje i objašnjavanje pojava.

Kao i svako drugo istraživanje i anketno je potrebno unaprijed i pažljivo planirati. Posebno se trebamo zadržati na preciznom definiranju cilja i problema istraživanja, definiranju prihvatljive razine greške te resursima koji su nam na raspolaganju (primarno vremenskim i materijalnim). Na osnovu ovih ulaznih parametara pravi se daljnji plan koji uključuje odluke o instrumentariju koji će biti primijenjen, postupku uzorkovanja te samoj veličini uzorka. Pristupi dizajniranju anketnog upitnika su detaljnije objašnjeni u poglavlju 3. u kontekstu mjerenja i dizajniranja mjernih instrumenata.

5.4.4. Načini unapređenja valjanosti anketnih istraživanja

Jedan od najčešće korištenih načina unapređenja valjanosti anketnih istraživanja je provođenje predistraživanja, tj. pilot istraživanja. Postupak provođenja pilot istraživanja podrazumijeva primjenu anketnog upitnika sa nešto većim brojem pitanja nego što je planirano da bude u finalnoj verziji, na manjem broju ispitanika. Pri tome se sa ispitanicima radi dubinska analiza razumijevanja pitanja i odgovora koje daju na način da ispitanici

izvještavaju o svom misaonom procesu. Na ovaj način dobijamo uvid u način razumijevanja pitanja od strane ispitanika te formuliranja odgovora. Ovaj korak je izuzetno važan jer nam daje uvid u to da li je pitanje formulirano na razumljiv i ispravan način te da li smo pitanjima obuhvatili sve aspekte koje treba ispitati.

Alternativa ili nadopuna ovog pristupa podrazumijeva konsultacije sa stručnjacima koji se bave područjem koje je predmet interesa. Profesionalci iz nekog područja nam mogu pomoći da provjerimo ispravnost terminologije i formulacija u pitanjima, dok s druge strane vjerovatno neće biti osjetljivi na jezične nepravilnosti i nejasnoće. Kako bi se prekontrolisao i jezik korišten prilikom formulacije pitanja, potrebno je konsultirati i stručnjake iz područja lingvistike i/ili prevođenja ukoliko koristimo kao osnovu upitnik na stranom jeziku. Prilikom prevođenja potrebno je dobro paziti na jasnoću jezičkih formulacija kako se ne bi gubio smisao u prevodu.

Dodatno, moguće je analizirati ponašanje ispitanika prilikom glavnog istraživanja, gdje bismo bili posebno osjetljivi na ponašanja koja odaju nerazumijevanje (upiti za pojašnjenjem, odustajanje od odgovaranja, nelogično ili neprimjereno odgovaranje). Problem sa ovim pristupom je u činjenici da ga je poprilično teško implementirati kada radimo istraživanje na velikom uzorku, kao i u tome što nam ne omogućava intervenciju na pitanjima na osnovu uočenih problema.

Jedan od načina testiranja valjanosti anketnog upitnika je poređenje odgovora na anketna pitanja i objektivnih mjera koje su dostupne o ponašanju koje pitamo. Ovaj postupak je teško provediv u većini slučajeva, posebno tamo gdje je anketiranje anonimno. Npr. ukoliko pitamo ispitanike da li su glasali na izborima, vjerovatnoća je da ćemo dobiti potvrde (socijalno poželjne odgovore). U teoriji, odgovor svakog ispitanika je moguće provjeriti ukrštanjem sa evidencijom na biračnom spisku. U praksi je ovaj postupak teško provesti, ipak, poređenje informacija datih u anketi na nivou grupe i onih koje (opet na nivou referentne grupe) imamo u biračkim spiskovima nam može dati generalnu informaciju o tome da li su ispitanici iskreni u odgovaranju.

5.4.5. Snaga i jasnoća stava

Kada smo do sada govorili o utjecaju načina formuliranja pitanja na iskazani stav, nismo adresirali faktor individualnih razlika. U suštini, do utjecaja na iskazani stav dolazi zbog, međudjelovanja faktora individualnih razlika i načina formuliranja pitanja. Drugačije rečeno, oblik pitanja, odabir riječi i red pitanja mogu utjecati na iskazani odgovor samo u slučaju ukoliko reaguju sa određenim osobinama ispitanika. Kada govorimo o individualnim karakteristikama, onda govorimo primarno o snazi i jasnoći stava.

Snaga stava se odnosi na intenzitet uvjerenja na kojem je temeljen stav. Npr. dvije osobe mogu imati naizgled isti stav kada govorimo o inkluziji u obrazovanju, gdje je jednoj osobi to dio nekog načenog principa koji primjenjuju u svemu, a drugoj jedno od najvažnijih životnih pitanja jer je se direktno tiče (npr. ima dijete sa teškoćama za koje smatra da treba biti uključeno u redovni obrazovni program). Jasni stavovi su oni koji su generalno neovisni o načinu mjerenja i generalno nisu samo formalni po svojoj prirodi nego suštinski.

Kao što možemo vidjeti, pojmovi snage i jasnoće stava se prepliću i teško ih je u potpunosti razdvojiti. Nama ovdje terminološko razlikovanje i nije primarni cilj. Primarni cilj je uvid u utjecaj koji snaga i jasnoća stava imaju na valjanost i pouzdanost anketnog istraživanja. Jasniji, snažniji stavovi su puno otporniji na vanjske utjecaje kao što je način formuliranja pitanja u odnosu na stavove koji su slabiji snagom i nejasniji.

6. Pregled statističkih postupaka koje često koristimo

Kako smo mogli vidjeti iz uvodnog poglavlja kada smo ponudili pregled vrste istraživanja i statističkog postupka koji je adekvatan alat u tom slučaju, kao i iz poglavlja o eksperimentu i kvaziekperimentu, osnova za zaključivanje o prihvatanju ili ne postavljene hipoteze je primjena nekog od statističkih postupaka koji nam omogućavaju poređenje mjerenja ili grupa. Prije odluke o tome koji ćemo statistički postupak za poređenje primijeniti moramo prvo provjeriti normalnost distribucije. Ukoliko je distribucija rezultata normalna, tada su nam dostupni statistički postupci koji pripadju parametrijskoj statistici. U slučaju distribucije koja statistički značajno odstupa od normalne, koristimo neparametrijske statističke postupke. Dalje, u zavisnosti od toga koji nivo mjerenja koristimo možemo porediti frekvencije ili aritmetričke sredine. Naravno da frekvencije i aritmetričke sredine nisu jedine vrijednosti koje možemo porediti, ali u istraživanjima upotrebljivosti se uspješno primjenjuju pa ćemo se zato fokusirati na njih u ovom tekstu.

Prije nego počnemo sa elaboracijom statističkih postupaka koje koristimo prilikom poređenja, potrebno je objasniti osnovne statističke koncepte deskriptivne i inferencijalne statistike.

6.1. Vrste varijabli

Da bismo na adekvatan čprocijenili nivo analize koji možemo uraditi i sofisticiranost zaključaka do kojih mozemo doći, prvo moramo poznavati vrstu podataka koje su nam na raspolaganju. U statistici razlikujemo nekoliko vrsta varijabli:

- Diskretne - postoji ograničen broj vrijednosti koje varijabla može imati, npr. spol, razred
- Kontinuirane moguća je bilo koja vrijednost u rasponu od minimalne do maksimalne vrijednosti - dob, masa, dužina i sl
- Kvantitativne predstavljaju numeričke rezultate bilo koje vrste mjerenja
- Kategorijalne - razvrstavamo pojavu u određeni broj kategorija.

6.2. Frekvencije

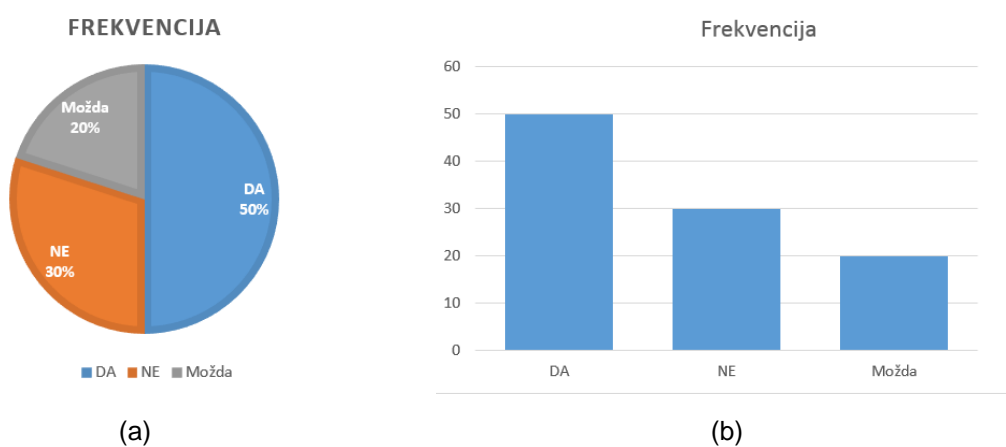
Jedan od osnovnih pojmova deskriptivne statistike i jedan je od alata koji se često koriste su frekvencije. Frekvencije nam govore o tome koliko se često neka vrijednost pojavljuje u vašem skupu podataka, te na osnovu toga dobijemo brz i djelotvoran uvid u to kako se nasi podaci dijele u kategorije. Na primjer, da bismo utvrdili da li je spol faktor za to kome je naš proizvod zanimljiv i percipiran kao koristan jedna od stvari koju bilo bi dobro znati je koliko je trenutnih korisnika muškaraca a koliko žena.

Frekvencije najčešće prikazujemo tabelarno, gdje nam jednostavna tabela daje puno informacija. Npr. Tabela frekvencija odgovora na pitanje: “Da li biste ovu aplikaciju preporučili prijatelju?”

Tabela 6.1 Primjer frekvencijske tabele

	DA	NE	Možda
Frekvencija	50	30	20

Ova jednostavna tabela za čitati i interpretirati nam daje mnogo informacija koje nam s obzirom na kontekst mogu biti vrlo zanimljive i/ili značajne. Naime, vidimo da od ukupno 100 ispitanika ($50 + 30 + 20$) njih pola izjavljuje da bi našu aplikaciju preporučilo svom prijatelju. Iz toga zaključujemo je da pola naših ispitanika ili pola korisnika naše aplikacije zadovoljno našim proizvodom. Pored 50 ispitanika koji su se izjasnili da bi našu aplikaciju preporučilo, njih 20 je dalo odgovor možda. U zavisnosti od toga šta nam je cilj prilikom razvijanja proizvoda, te cilj istraživanja, ovih 20 ispitanika (i njihovi istomišljenici u populaciji od interesa - pojedinci koji koriste naš proizvod) su nam vrlo zanimljivi. Ispitanici koji su neodlučni da li da preporuče našu aplikaciju ili ne nam mogu dati korisne informacije o prednostima i nedostacima i samim time i upute na koji način proizvod unaprijediti. Ovi ispitanici mogu biti predmetom dublje analize tj. mogu biti ispitanici u nastavku istraživanja kako bismo dobili dodatne informacije o njihovom vidjenju proizvoda te na koji način ga unaprijediti da zadovolji njihove kriterije. Na kraju, 30 ispitanika (što je 30% našeg uzorka i ukoliko je uzorak reprezentativan, možemo reći i 30% svih korisnika koji su koristili naš proizvod, ga ne bi preporučili prijatelju iz čega možemo zaključiti da nisu zadovoljni onim što smo ponudili. U zavisnosti od toga šta nam je bio cilj kad smo razvijali proizvod i od toga kolika nam je ciljna populacija ova informacija može biti ili ne mora osnova za zabrinutost. Isti primjer je iskorišten za ilustraciju grafičke vizualizacije frekvencija, korištenjem stupičastog i pita dijagrama prikazanih na slici 6.1.



Slika 6.1 Primjeri vizualizacije frekvencija korištenjem (a) pita i (b) stupičastog dijagrama

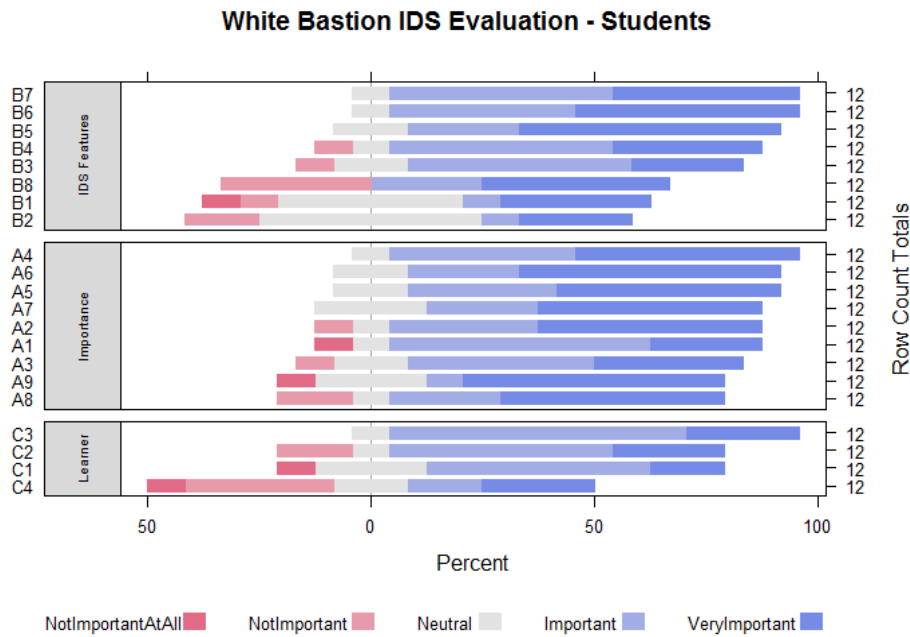
Postupak za kreiranje tabele frekvencija je jednostavan. Nakon što se odredi skup relevantnih kategorija za analizu (u ovom slučaju su tri alternative koje su date kao mogući odgovor na pitanje) za skup prikupljenih podataka, jednostavno je pitanje brojanja određenih vrijednosti koje spadaju u definirane kategorije (napomena: ovdje koristimo termin kategorije zbog toga što prilikom kvalitativne analize prvo definiramo kategorije u koje ćemo svrstavati sve odgovore pa pristupamo brojanju). Prilikom kvalitativne analize gdje definiramo kategorije u koje ćemo svrstavati odgovore, postupak uglavnom radimo mi a ne koristimo neki od programa za obradu u ovoj fazi. Nakon što smo odgovore razvrstali u kategorije, podatke unosimo u neki od programa za statističku obradu i daljnja analiza je moguća. Ipak, potrebno je napomenuti da postoje programska rješenja za kvalitativnu analizu koja je opravdano koristiti ukoliko imamo izuzetno veliki uzorak pa nam preciznost i dubina analize nisu primarni na nivou svakog ispitanika.

Primjer istraživanja u kojem se frekvencije odgovora ispitanika prezentiraju i tabelarno i grafički preuzet je iz rada Rizvić i sur. (2019) i tabelarni je prikazan u tabeli 6.2., a grafički na slici 6.2. U ovom primjeru koji se bavi evaluacijom korštenja interaktivnog digitalnog pripovjedanja o kulturnom nasljeđu (primjer Bijele Tabije) u nastavi historije, ispitanici su podijeljeni u grupe s obzirom na ulogu u nastavnom procesu: nastavnici (N) i učenici (S), kao i posebna grupa ispitanika (D) koja nije u vrijeme ispitivanja uključena u nastavni proces. Ovaj primjer uključuje više grupa pitanja, i u tabeli 6.2. su prikazani rezultati jedne grupe pitanja (A), dok su na slici 6.2. grafički vizualizirani odgovori iz sve tri grupe pitanja (A, B i C) za grupu ispitanika studenti.

Tabela 6.2. Primjer frekvencijske tabele iz realnog istraživanja

Pitanje	Uopšte nije važno			Nije važno			Nemam mišljenje			Važno je			Veoma je važno		
	N	S	D	N	S	D	N	S	D	N	S	D	N	S	D
A1	0	1	1	0	0	0	0	1	0	5	7	5	13	3	13
A2	0	0	1	0	1	0	0	1	0	6	4	6	12	6	12
A3	0	0	1	0	1	0	2	2	2	11	5	5	5	4	11
A4	0	0	1	0	0	0	0	1	1	7	5	5	11	6	12
A5	0	0	1	0	0	0	0	2	1	9	4	2	9	6	15
A6	0	0	1	0	0	0	0	2	1	10	3	3	8	7	14
A7	0	0	1	0	0	0	2	3	1	5	3	6	11	6	11
A8	0	0	1	0	2	0	1	1	0	9	3	4	8	6	14
A9	0	1	1	1	0	0	0	3	0	8	1	2	9	7	16

Za ilustraciju frekvencija odgovora na Likertovoj skali mjerenja korišten je kompozitni stupičasti dijagram (eng. *stacked bar*) gdje svaki stupac odgovara jednom pitanju, a različiti odgovori su kodirani različitim bojama. Primjer je isprogramiran u jeziku R i to primjenom divergentnih stupaca koji su centrirani na dijelu stupca koji odgovara neutralnom odgovoru. Ovaj način vizualizacije su predložili Heiberger i Robbins (2014).



Slika 6.2. Primjer vizualizacije frekvencija odgovora na Likertovoj skali korištenjem dijagrama divergentnih stubaca

6.3. Mjere centralne tendencije i mjere varijabiliteta

Dobivanje uvida u distribuciju prikupljenih podataka je važan inicijalni korak u statističkoj obradi podataka važan za izbor statističkih metoda koje ćemo dalje koristiti. Potrebno je da izračunamo ključne numeričke vrijednosti o podacima koji će nas informisati o o distribuciji podataka, i da vizualizacijom steknemo uvid u oblik distribucije. Ključne numeričke karakteristike distribucije su centar podataka i mjera njihovog rasipanja, odnosno varijabiliteta.

6.3.1. Mjere centralne tendencije

Mjere centralne tendencije nam daju osnovne informacije o podacima smo prikupili, tako da možemo bolje razumjeti svoje korisnike, kvalitetnije usmjeriti prodaju i marketinške napore. Mjere centralne tendencije uključuju aritmetričku sredinu, medijanu i mod. Ovdje ćemo se fokusirati na aritmetričku sredinu, kao najinformativniju i najčešće korištenu mjeru centralne tendencije. Napominjemo da se ove kvalifikacije aritmetričke sredine odnose na situacije kada je naša distribucija rezultata odgovarajuća za računanje aritmetričke sredine. Minimalni uslov za računanje aritmetričke sredine je simetrična distribucija. Idealno, želimo normalnu distribuciju. Pored toga, dobijene rezultate treba pogledati u njihovom sirovom obliku pa vidjeti da li postoje strčeće vrijednosti (eng. *outlier*) koje remete izgled distribucije. Te strčeće vrijednosti su nekada očito rezultat greške u procesu mjerenja, a nekada su odraz stvarnog stanja. U prvom slučaju je opravdano eliminirati taj rezultat iz skupa podataka, dok u drugom moramo uzeti više faktora u obzir prije odluke o eliminiranju rezultata. Ukoliko imamo mali uzorak veća je vjerovatnoća strčećih rezultata i samim tim je odabir aritmetričke sredine kao mjere centralne tendencije problematičniji. U većem uzorku ekstremni rezultati manje utječu na izgled distribucije i ne pomjeraju aritmetričku sredinu značajno što nas čini

sigurnijim u odabir aritmetričke sredine kao mjere centralne tendencije i posljedično sigurnijim u svoje zaključke.

Mjere centralne tendencije nam omogućavaju da pronađemo sredinu skupa podataka. Ova sredina—bilo da je u pitanju aritmetrička, medijana ili mod—je brz i lak način da se skup podataka sumira u jedan broj. Poznavanje aritmetričke sredine, medijane ili moda skupa podataka može pomoći da brzo i jednostavno dobijemo uvid u iskustvo naših korisnika ili da usmjerimo napore za unapređenje proizvoda.

6.3.1.1. Aritmetrička sredina

Aritmetrička sredina skupa podataka je prosječna vrijednost tih podataka. Za podatke prikupljene tokom vremena, aritmetričku sredinu je najbolje koristiti kada se podaci bilježe redovno – na primjer svaki dan, sedmica ili mjesec – tako da možete uzeti u obzir mjere poput dnevnih prosjeka (npr. prosječan broj posjeta web stranici dnevno). Za podatke koji karakteriziraju grupisane vrijednosti, kao što su različite dobi unutar grupe korisnika, aritmetrička sredina se najbolje koristi kada nema strčućih rezultata tkz. outliera (vrijednosti koje su neobično visoke ili neobično niske u odnosu na druge vrijednosti u grupi).

Za izračunavanje aritmetričke sredine, saberete svaku vrijednost u skupu podataka i podijelite taj zbroj s brojem vrijednosti:

$$\text{Aritmetička sredina} = \frac{\sum_{i=1}^n x_i}{n}$$

Aritmetrička sredina, kada se računa na skupu uzoraka označava se kao \bar{x} , a kada se računa na svim obezvrvacijama populacije označava se μ .

6.3.1.2. Medijana

Medijana je vrsta mjere centra skupa podataka, koja i doslovno ukazuje na centar podataka, dijeli skup podataka na dva podskupa sa jednakim brojem elemenata. Kod skupa podataka koji imaju neparan broj elemenata, medijana ima vrijednost jednaku podatku u centru kada su poredani od najmanjeg do najvećeg. Kada skup podataka ima paran broj elemenata onda medijana ima vrijednost aritmetičke sredine dva podatka u centru.

Medijana je manje osjetljiva na strčće vrijednosti od aritmetičke sredine, što je čini pogodnom za mjeru centralne tendencije za nepravilne asimetrične distribucije podataka.

6.3.1.2. Mod

Mod je mjera centralne tendencije koja nam ukazuje na mjesto gomilanja vrijednosti. Mod je vrijednost koje je najčešća u skupu podataka. Skup podataka može imati više vrijednosti koje predstvaljaju mod. Distribucija podataka koja ima samo jedan mod se naziva unimodalna. Distribucija podataka sa dva moda se naziva bimodalna, a općenito, distribucije s više od jednog moda se nazivaju multimodalne. Mod je mjera centralne tendencije koja se može odrediti i za kategoričke i za numeričke podatke.

6.3.2. Mjere varijabiliteta

Mjere varijabiliteta mogu biti važan alat za razumijevanje podataka o upotrebljivosti. Ove mjere su jedan od osnovnih alata da opišemo varijansu unutar skupa podataka. U ovom kontekstu, varijansa se odnosi na broj koji izražava koliko su “raštrkani” naši podaci od srednje vrednosti. Dakle, mjere varijabiliteta su alati koje koristimo da opišemo koliko vrijednosti unutar našeg skupa podataka odstupaju od srednje vrijednosti tog skupa.

U statističkoj analizi podataka, varijansa se koristi na mnogo različitih načina. To je važan dio analize rizika jer ako postoji visok stepen varijanse u našim podacima, ne možete precizno predviđati ishode iz tih podataka. Shodno tome, radnje zasnovane na tim predviđanjima predstavljaju veći rizik da budu pogrešne. Varijansa je važan pokazatelj pouzdanosti podataka. Općenito, pomaže nam da odredimo koliko možemo vjerovati onome što se čini da naši podaci govore. Osim toga, varijansa može pomoći kreatorima proizvoda kao i onima koji proizvod plasiraju na tržište da donesu odluke o tome ko bi mogao biti razumna ciljna publika, ili da procijene da li je ciljana kampanja bila uspješna u doseganju svoje publike. Jednostavno rečeno, mjere varijabiliteta mogu pomoći da usmjerimo svoje napore kako poboljšali vjerovatnoću da postignemo željene ciljeve.

6.3.2.1. Varijansa i standardna devijacija

Standardna devijacija je najpreciznija mjera varijabiliteta rezultata i uvijek je računamo sa aritmetičkom sredinom. Ako za računanje standardne devijacije koristimo aritmetičku sredinu uzoraka \bar{x} , označavamo je s ; a ako je računamo za populaciju i na bazi aritmetičke sredine populacije μ , označavamo je σ . Varijansa je jednaka kvadratu standardne devijacije, i opet u zavisnosti da li je u pitanju varijansa uzorka ili populacije označavamo je s^2 , odnosno σ^2 . Formule za računanje standardne devijacije za uzorak i populaciju su:

$$s = \sqrt{\frac{\sum_i^n (x_i - \bar{x})^2}{n - 1}}$$

$$\sigma = \sqrt{\frac{\sum_i^n (x_i - \mu)^2}{n}}$$

Možemo reci da je standardna devijacija najpreciznija mjera varijabiliteta. Standardna devijacija nam daje određeni raspon vrijednosti koji obuhvata određeni postotak našeg skupa podataka. Na primjer, standardna devijacija se može koristiti za analizu skupa podataka o prodaji i reći će nam da, na osnovu prošlih marketinških napora, 68% kupaca vjerovatno potrošiti između 18 i 21 BAM. Standardna devijacija to čini razbijanjem kompletnog skupa podataka u standardizirane blokove, na osnovu udaljenosti od srednje vrijednosti.

6.3.2.2. Opseg i kvantili

Opseg ili raspon je mjera varijabiliteta koja opisuje koliko su rašireni podaci. Opseg je najjednostavnija mjera varijacije i predstavlja razliku između najveće i najmanje vrijednosti podataka.

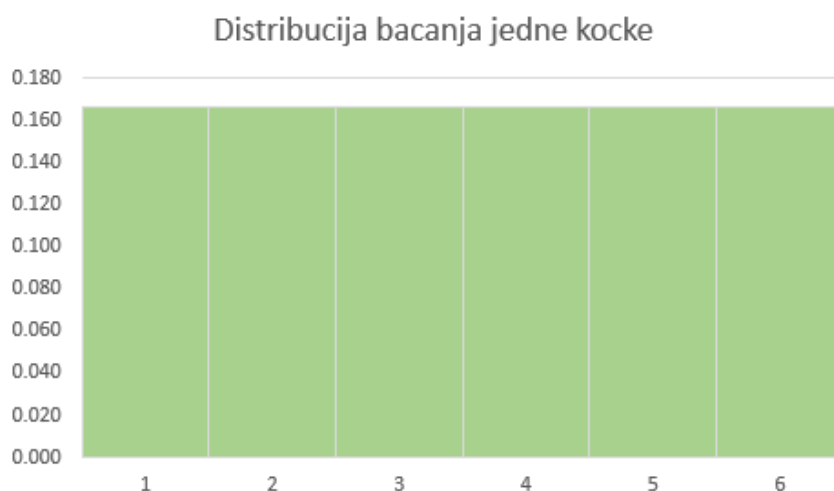
Kvantili su mjera varijabiliteta koja također opisuje koliko su podaci rašireni. Vrste kvantila koji se često koriste u statistici su kvartili i percentili. Kvartili (Q0,Q1,Q2,Q3,Q4) su vrijednosti koje dijele skup podataka na četiri jednaka dijela. Između Q0 i Q1 nalazi se 25% najnižih vrijednosti u podacima, između Q1 i Q2 je sljedećih 25%, a između Q0 i Q2 se nalazi pola vrijednosti podataka. Očigledno je da je Q0 najmanja vrijednost u podacima – minimum, Q2 je srednja vrijednost ili medijana, a Q4 je najveća vrijednost u podacima - maksimum. Razlika između Q4 i Q0 je opseg (raspon), a razlika između Q3 i Q1 se naziva interkvartilni raspon (IQR) i obuhvata polovinu vrijednosti iz skupa podataka i to „srednju polovinu“ podataka, koncentrisanu oko srednje vrijednosti.

Percentili su vrijednosti koje dijele podatke na 100 jednakih dijelova. Na primjer, 95. percentil odvaja najnižih 95% vrijednosti od gornjih 5%. Prvi kvartil Q1 je jednak 25. percentilu (P25%), medijana je i Q2 i 50. percentil (P50%) i tako dalje.

6.4. Raspodjele podataka

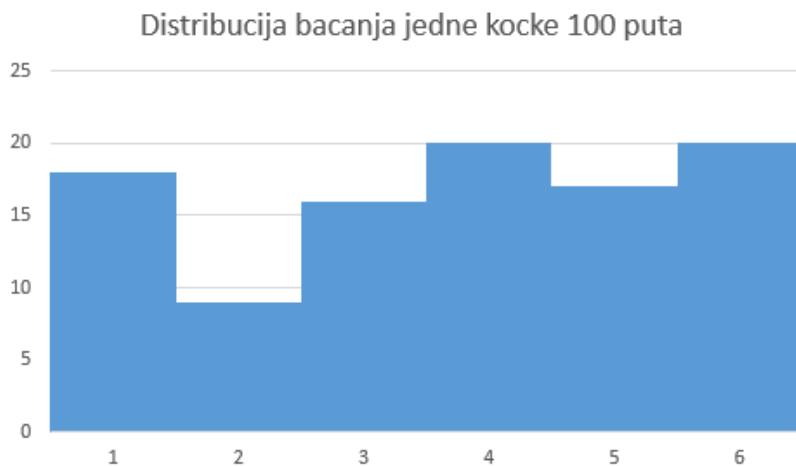
Raspodjele (distribucije) podataka pokazuju vjerovatnoću mogućih ishoda unutar nekog skupa podataka. Distribucije predstavljamo grafički krivuljom distribucije. Precizan matematski termin je funkcija gustoće vjerovatnoće – PDF (eng. *Probability Density Function*). Za kontinuirane slučajne varijable ovo je krivulja koja nam daje mogućnost korisne geometrijske interpretacija vjerovatnoće da naša varijabla ima vrijednost u određenom rasponu vrijednosti. U slučaju diskretne slučajne varijable koja ima više mogućih vrijednosti i svaka ima različitu moguću vjerovatnoću, funkcija gustoće vjerovatnoće je jednostavno skup svih vjerovatnoća za sve različite ishode.

Ovo ćemo ilustrovati primjerom bacanja jedne kocke. Jasno je da je varijabla koju posmatramo diskretna, i zapravo proces mjerenja je brojanje. Kod bacanje jedne kocke imamo šest različitih ishoda koji imaju jednaku vjerovatnoću i to jednaku jednoj šestini. Jasno je da je suma vjerovatnoća svih mogućih ishoda 1, jer neki ishod se mora dogoditi. Primjer funkcije gustoće vjerovatnoće za bacanje jedne kocke je dat na slici 6.3. Prikazana distribucija se naziva uniformna distribucija baš zbog jednake vjerovatnoće svih ishoda.



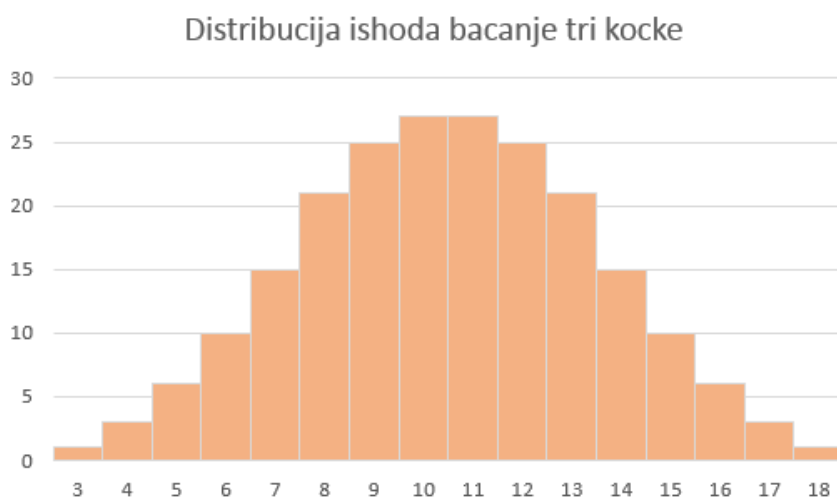
Slika 6.3. Primjer funkcije gustoće vjerovatnoće za bacanje jedne kocke

U slučaju da napravimo eksperiment i bilježimo ishode bacanja jedne kocke dobit ćemo distribuciju stvarnog skupa podataka. Ovako jednostavni eksperimenti se lako mogu simulirati računarski, pa ćemo rezultate eksperimenta bacanja kocke 100 puta ilustrirati na slici 6.4. Ova distribucija dobijena eksperimentom nije jednaka funkciji gustoće vjerovatnoće, ali sa povećanjem broja bacanja približili bi se i više uniformnoj raspodjeli.



Slika 6.4. Primjer distribucije vrijednosti bacanja jedne kocke za eksperiment od 100 bacanja

Drugi primjer funkcije gustoće vjerovatnoće ćemo dati za slučaj bacanja tri kocke, a varijabla koju posmatramo, ishod bacanja, bit će zbir vrijednosti na sve tri kocke. Ovaj primjer je određen sa šest neovisnih mogućih ishoda za jednu kocku, koji za tri kocke daju 216 mogućih ishoda. Vjerovatnoća ishoda više nije jednaka, jer ishodi jednaki zbiru 3 ili 18 su vezani za samo jednu kombinaciju pa je njihova vjerovatnoća jednaka $1/216$. Navešćemo primjer za ishod jednak zbiru 5 kojem odgovaraju sljedeće kombinacije: „1, 1, 3“, „1, 3, 1“, „3, 1, 1“, „1, 2, 2“, „2, 1, 2“ i „2, 2, 1“. Najveću vjerovatnoću imaju ishodi jednaki 10 i 11, jer su rezultat više mogućih kombinacija. Primjer funkcije gustoće vjerovatnoće za bacanje tri kocke je dat na slici 6.5.



Slika 6.5. Primjer funkcije gustoće vjerovatnoće za bacanje tri kocke

Govorili smo o konceptu varijanse kao načinu da se odredi koliko je skup podataka raspršen od svoje srednje vrijednosti. Ovaj koncept je snažno povezan sa konceptom

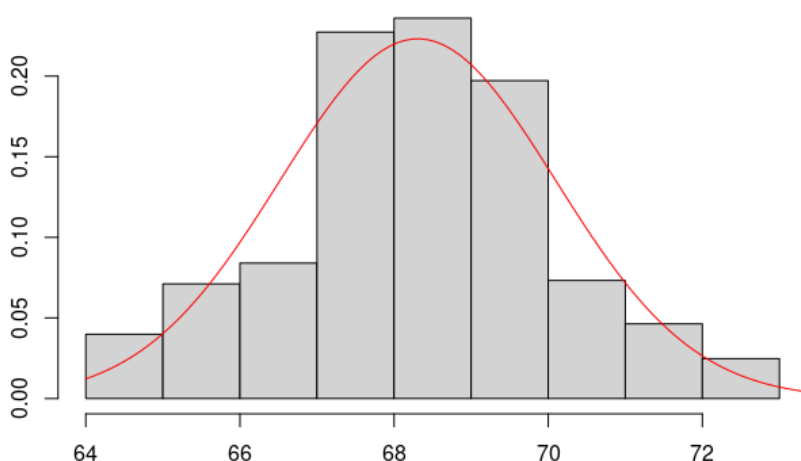
distribucije. Varijansa mjeri koliko je naša distribucija raširena, dok distribucija izražava specifičnu vjerovatnoću pojave bilo koje vrijednosti. Ako imate veliku varijansu, imat ćete nisku i široku krivu distribucije, jer je vjerovatnoća da vrijednost bude oko srednje vrijednosti samo malo veća od vjerovatnoće da bude bilo gdje drugdje. Ako imate malu varijansu, imat ćete visoku, usku distribuciju, jer postoji velika vjerovatnoća da će bilo koja vrijednost biti blizu srednje vrijednosti. Pogledajmo bliže kako vizualizirati distribuciju skupa podataka.

6.4.1. Histogrami

Distribucije vrijednosti se mogu vizualizirati pomoću grafikona koji se naziva histogram. Histogrami su vrsta stupičastog grafikona, gdje je svaki stupac definisan širinom tj. rasponom vrijednosti na x-osi i visinom - brojem podataka koji spadaju u taj raspon. Ovi rasponi se mogu nazvati i razredima, ali su kod svih stupaca isti. Visina svakog stupca u histogramu će biti određena brojem ponavljanja vrijednosti podataka koje spadaju u taj razred. Dakle, razredi u vašem histogramu predstavljaju raspon vrijednosti na grafikonu, a visine stupaca su zbir frekvencija za svaku vrijednost u razredu.

Primjer histograma je dat na slici 6.3. gdje smo ilustrovali rezultate eksperimenta bacanja jedne kocke sto puta. Razredi stupaca u ovom slučaju obuhvataju samo jednu vrijednost na x-osi, što je u ovom primjeru bilo moguće jer imamo mali broj različitih ishoda – šest, pa nije bilo potrebno da ih objedinjavamo. Histogrami se mogu praviti i za kontinuirane varijable i tada razredi obuhvataju interval vrijednosti. Primjer su visina, težina, i sl.

Histogram se može koristiti za poređenje distribucije podataka sa teorijskim modelom, kao što je normalna distribucija. Zato je korisno kod vizualizacije histograma uz stupičasti graf nacrtati i teoretsku krivulju raspodjele koja odgovara skupu podataka – ima istu srednju vrijednost i standardnu devijaciju. Na slici 6.6. je dat histogram raspodjele visina roditelja iz Galtonovog seminalnog istraživanja (Galton, 1886) i odgovarajuća krivulja normalne raspodjele.



Slika 6.6. Primjer histograma koji odgovara normalnoj distribuciji

Evo histograma koji odgovara raspodjeli datoj krivuljom normalne distribucije, što je vrlo čest oblik distribucije koji smo već spominjali i koji ćemo detaljnije pogledati u nastavku.

Krivulja distribucije za ovaj skup podataka ima oblik zvona, dakle zakrivljene linije i simetrična je što je, tipično za normalnu distribuciju.

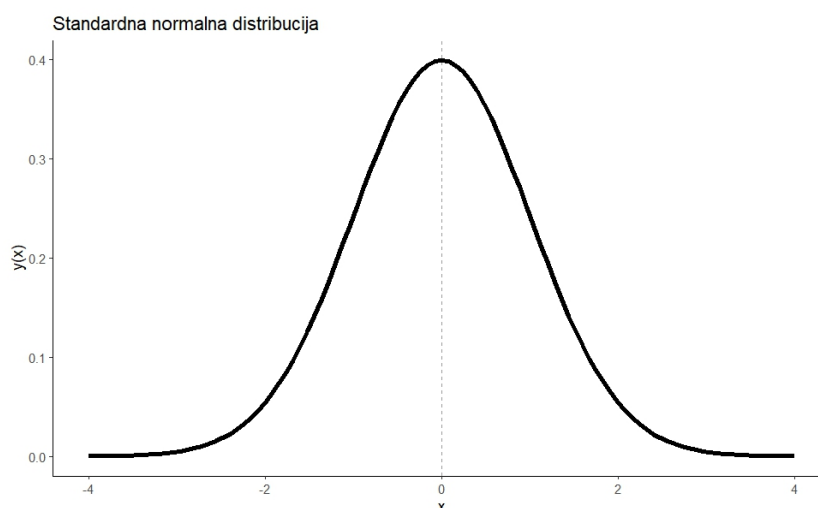
Teoretski, distribucija može imati bilo koji oblik, ovisno o podacima, ali postoje četiri krivulje distribucije na koje ćete najvjerojatnije naići tj. koje se najčešće javljaju: normalna, ujednačena, Poissonova i eksponencijalna. U nastavku su opisane i ilustrovane ove krivulje.

6.4.1.1. Normalna distribucija

Normalna distribucija je najčešći oblik distribucije jer se stalno javlja u prirodi. Ponekad se normalna distribucija naziva "zvonasta kriva", jer je oblikovana kao zvono i ima dvije važne karakteristike koje je odvajaju od ostalih:

1. Srednja vrijednost, medijan i mod imaju istu vrijednost u normalnoj distribuciji.
2. Takođe, kriva koja razlikuje normalnu distribuciju je simetrična, tako da kriva ispod srednje vrednosti izgleda isto kao i kriva iznad srednje vrednosti.

Na primjer, recimo da gledate koliko kupac troši po kupovini i otkrijete da njihova potrošnja spada u normalnu distribuciju. To ukazuje da će većina vaše prodaje biti oko srednje vrijednosti, s jednakim brojem ljudi koji će trošiti manje i više od prosjeka.



Slika 6.7. Kriva normalne distribucije

Normalna distribucija se matematski opisuje sljedećom formulom:

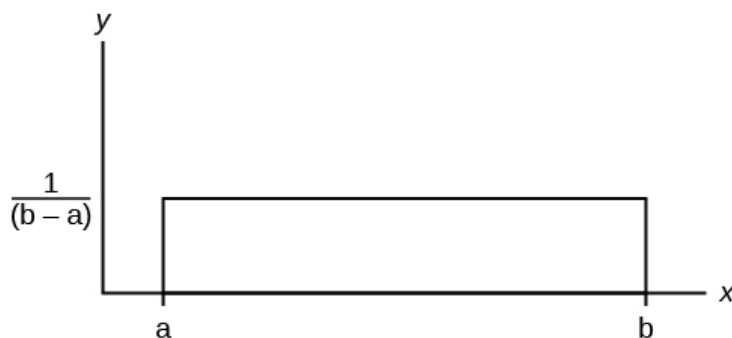
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Prethodna formula nam ukazuje na uticaj koji srednja vrijednost (μ) i varijansa, odnosno standardna devijacija (σ), imaju na oblik ove krive. Razlike u oblicima normalne distribucije i odstupanja će biti opisani u nastavku u poglavlju o normalizaciji podataka.

6.4.1.2. Uniformna distribucija

Uniformna distribucija, koja se ponekad naziva i pravokutna distribucija, je ravna linija koja se javlja kada se sve vrijednosti u vašem skupu podataka javljaju s istom frekvencijom. Svaka vrijednost ima jednake šanse da se dogodi. Na primjer, ovu vrstu distribucije možemo vidjeti ako imamo tri trgovine na sličnim lokacijama i otkrijemo da sve imaju otprilike isti

broj prodaja nakon određene reklamne kampanje. U tom slučaju distribucija bi izgledala otprilike ovako:



Slika 6.8. Uniformna distribucija

Kako površina ispod krivulje distribucije treba biti jednaka 1, ovaj primjer poopštava uniformnu distribuciju, tako da definiše vrijednost na y-osi u zavisnosti od opsega na x-osi (b-a), pa je i analitički oblik ove funkcije dat:

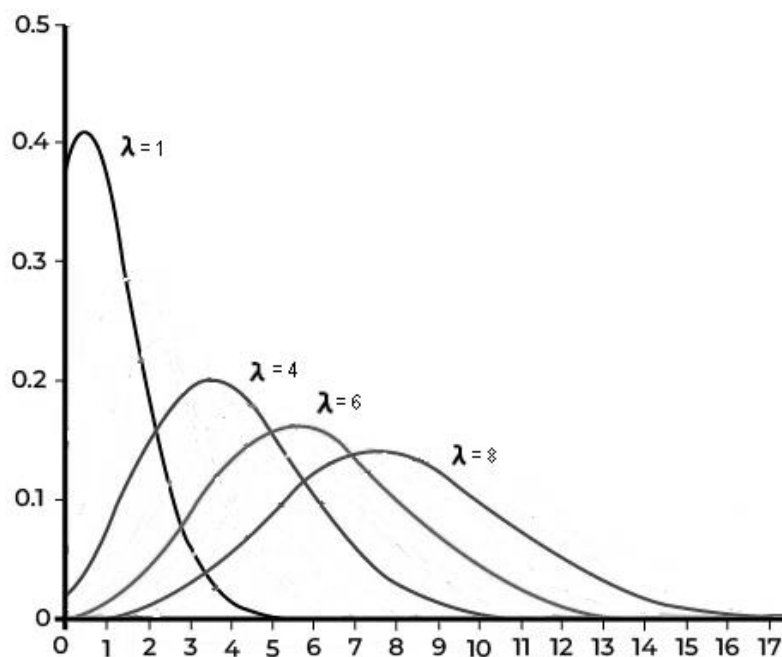
$$f(x) = \frac{1}{b-a}$$

Na primjeru bacanja jedne kocke koji je dat na slici 6.2. može se pročitati da je visina pravougaonika 0,167, što pomnoženo sa 6 ishoda daje vrijednost površine 1.

6.4.1.3. Poissonova distribucija

Poissonova distribucija se razlikuje od ostalih i predstavlja vjerovatnoću broja pojave slučajnih događaja unutar određenog vremenskog intervala. Na primjer, ako želimo znati vjerovatnoću prodaje, na osnovu toga koliko ljudi klikne na pop-up oglas na našoj web stranici, htjet ćemo pogledati Poissonovu distribuciju za naše podatke o klikovima po prodaji. Očigledno je naša varijabla diskretna.

Kod Poissonove distribucije, srednja vrijednost i varijansa su jednake, jer je srednja stopa pojave događaja (λ) jednaka širini raspodjele, odnosno varijabilnosti. Ovo je specifično za Poissonovu distribuciju i predstavlja njenu ključnu osobinu. Vizualizacija za Poissonovu distribuciju može imati različite oblike u zavisnosti od srednje vrijednosti pojave događaja, odnosno stope λ . Evo primjera kako bi takva distribucija mogla izgledati, za nekoliko različitih srednjih vrijednosti λ :

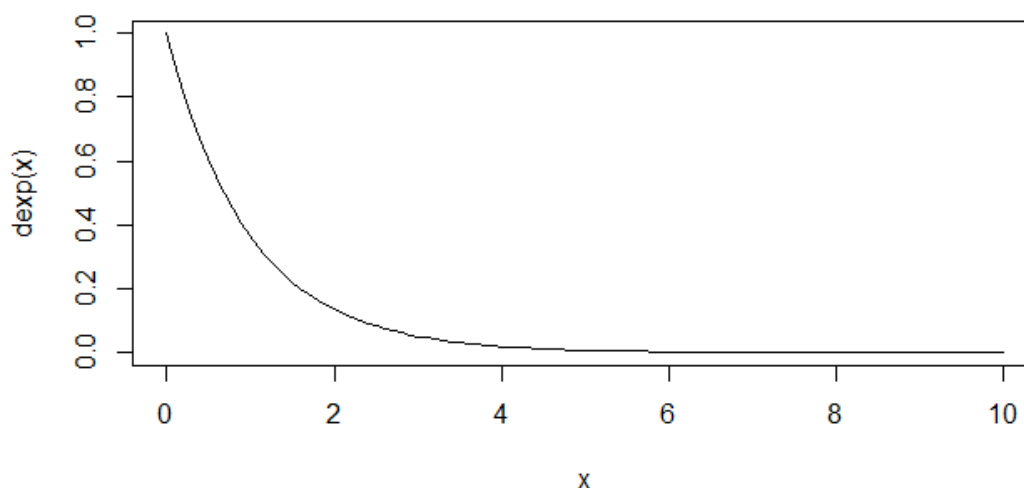


Slika 6.9. Poissonova distribucija

6.4.1.4. Eksponencijalna distribucija

Eksponencijalna distribucija je vezana za Poissonovu distribuciju, jer se isto koristi za podatke snimljene kroz vrijeme, samo sada posmatramo vjerovatnoću trajanja vremena između pojave dva slučajna događaja. Oblik eksponencijalne distribucije pokazat će oštru krivulju koja može biti orijentirana u bilo kojem smjeru: može se zakriviti prema gore kroz vrijeme ili se može zakriviti prema dolje kroz vrijeme. Eksponencijalna distribucija je inverzna funkcija Poissonovoj distribuciji pojave događaja, jer što je događaja više, vrijeme između njih je manje, i obrnuto..

Na primjer, ako gledate koliko minuta je osoba na web stranici kako biste odredili kolika je vjerovatnoća da će obaviti kupovinu, možda ćete otkriti da prvih nekoliko minuta postoji naglo povećanje vjerovatnoće da će kupac nešto i kupiti, ali se to izjednačava kako vrijeme odmiče. Ovo odražava jasnu eksponencijalnu distribuciju i sugerira kako može postojati velika razlika između ljudi koji ostaju na našoj web stranici pet minuta umjesto jedne minute, ali vrlo mala razlika između onih koji ostaju na web stranici jedan sat ili 45 minuta.



Slika 6.10. Eksponencijalna distribucija

6.4.2. Normalizacija podataka

U prethodnim poglavljima objašnjeno je koliko statističkih analiza zahtijeva da imamo normalnu distribuciju. U našoj raspravi iznad, opisali smo četiri uobičajena tipa distribucija u idealnim terminima, za podatke koji savršeno spadaju u te krivulje distribucije. U stvarnom svijetu, podaci se često ne uklapaju uvijek tako savršeno. U onim trenucima kada naša analiza ovisi o normalnoj distribuciji, možda ćemo otkriti da je distribucija naših podataka ponekad blizu idealnoj normalnoj distribuciji, ali ne baš u potpunosti takva. Postoje dva uobičajena načina na koji naši podaci mogu odstupati od savršeno normalne distribucije: mogu biti iskrivljeni ili možemo imati ekstremne vrijednosti.

Prisjetimo se glavnih karakteristika idealne krive normalne distribucije. Rekli smo da normalna distribucija treba da bude simetrična i zvonasta, tako da kriva ispred srednje vrednosti izgleda isto kao i kriva iza srednje vrednosti. Iskrivljenost (eng. *skewnes*) i spljoštenost (eng. *kurtosis*) samo opisuju dva načina na koja distribucije u stvarnom svijetu ne budu simetrične i zvonastog oblika.

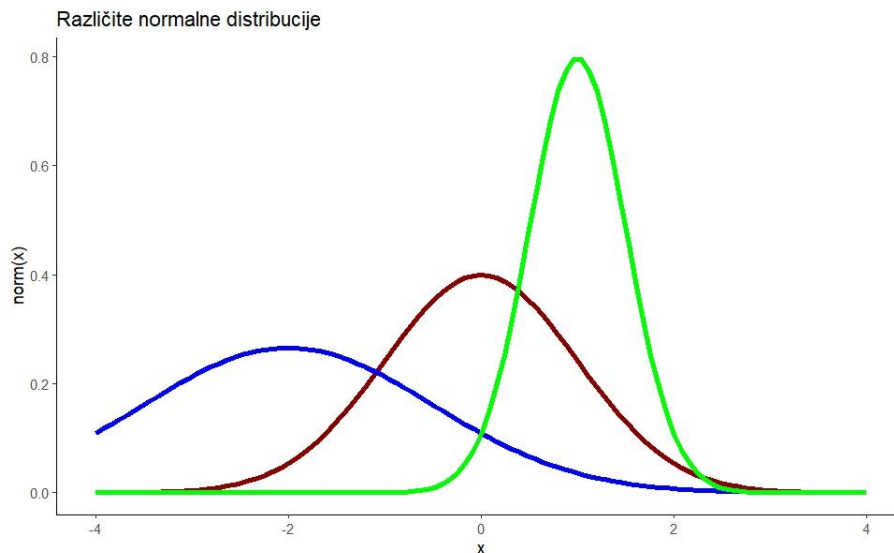
6.4.2.1. Iskrivljenost

Dok će simetrična distribucija biti ista na svakoj strani srednje vrijednosti, iskrivljena distribucija će "povući" na jednu ili drugu stranu srednje vrijednosti. Ako naši podaci imaju pozitivnu kosinu, većina naših podataka je na lijevoj strani i nešto vuče distribuciju na pozitivnu stranu, tako da na kraju imate dugi krak na desnoj strani. Ukoliko podaci imaju negativnu iskrivljenost, većina vaših podataka je udesno s dugim repom koji se proteže u negativnom smjeru ili lijevo.

6.4.2.2. Spljoštenost

Dok iskrivljenost opisuje kako se podaci "povlače" ulijevo ili udesno u poređenju sa savršenom simetrijom normalne distribucije, spljoštenost ili kurtosis opisuje kako se podaci mogu povući gore ili dolje, u poređenju sa savršenim oblikom zvona normalne distribucije. Moguće je da su na podaci takvi da distribucija ima visok, uski oblik, tako da krivulja

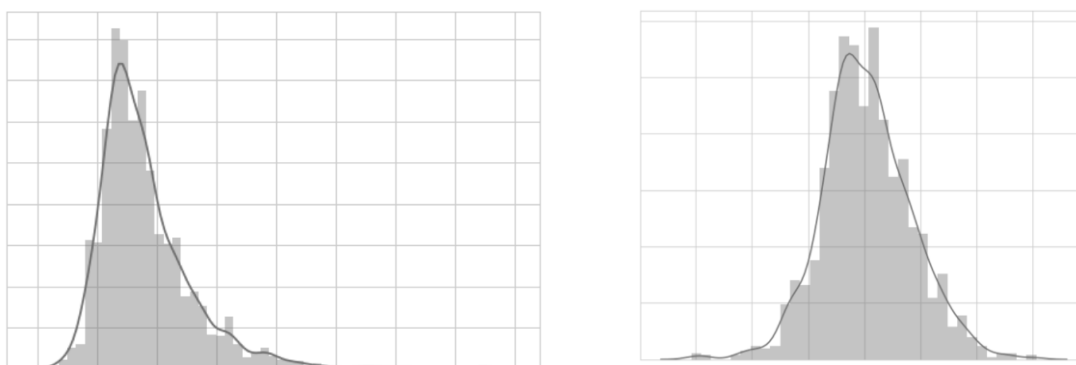
distribucije završava sa oštrijim vrhom u sredini. To bi značilo da je većina naših podataka čvršće grupisana oko srednje vrijednosti našeg skupa podataka (mala varijansa podataka). Negativan kurtosis povlači distribuciju prema dolje u sredini, tako da se kriva izravna u nizak, širok oblik. Ovakav oblik distribucije bi ukazivao na to da su naši podaci više „rasprostranjeni“ od srednje vrijednosti vašeg skupa podataka (velika varijansa podataka).



Slika 6.11. Krive normalne raspodjele različite spljoštenosti i iskrivljenosti

6.4.2.3. Transformacije i odstupanja

U prethodnim poglavljima smo objasnili da ukoliko naši podaci pokazuju previše iskrivljenosti ili ekstremnih rezultata, nećemo moći sprovesti statističke analize koje zahtijevaju normalnu distribuciju. Međutim, moguće je primijeniti tkz. transformacije na podatke kako bi se ispravila iskrivljena distribucija. Transformacije su načini na koje možemo pokušati "normalizirati" svoje podatke ili ukloniti iskrivljenost. Da bismo transformirali podatke, primjenjujemo jednostavnu matematičku jednačinu na svaku vrijednost u našem skupu podataka. Nakon toga transformirani podaci bi trebali pokazati novu distribuciju koja uklanja iskrivljenost originalnih podataka. Nakon toga možemo uraditi analizu koja zahtijeva normalnu distribuciju.



Slika 6.12. iskrivljena distribucija vremena izvršenja i distribucija vrijednosti dobijenih logaritamskom transformacijom

Primjer je distribucija vremena izvršenja koja je tipično iskrivljena na desno, slika 6.12. lijevo. Nekoliko vrijednosti za duže vrijeme izvršenja (za pojedine korisnike) ne samo da utiče na grafički prikaz distribucije, već značajno utiče na srednju vrijednost kao mjeru centralne tendencije. U ovom slučaju medijan je bolja mjera za ocjenu vremena izvršavanja ili mod kao najčešća vrijednost ove metrike. Logaritamskom transformacijom izmjerenih vrijednosti dobije se raspodjela prikazana na slici 6.12. desno.

6.5. Kompariranje dvije aritmetričke sredine (t-test)

Vrlo često, bilo korištenjem eksperimenta, anketnog istraživanja ili nekog drugog postupka prikupljanja podataka, tražimo odgovor na pitanje postojanja razlika između grupa korisnika. Npr. dvjema grupama korisnika ponudimo dvije aplikacije koje omogućavaju obavljanje istih zadataka i prikupljamo podatke o njihovom iskustvu i vremenu potrebnom da se zadatak izvrši. S obzirom na to da nas zanima donošenje generalnog zaključka (koja je aplikacija bolja za populaciju od interesa), moramo analizirati grupne podatke. Nažalost nije dovoljno samo izmjeriti aritmetričku sredinu i standardnu devijaciju i donijeti zaključak o postojanju razlike. Izmjerene razlike mogu biti rezultat slučaja ili neke nesistemske greške u mjerenju i potrebno je provjeriti da li su dovoljno velike (s obzirom na način mjerenja i veličinu uzorka) da smo sigurni da stvarno postoje.

Kako bismo došli do znanstvenog odgovora na pitanje postojanja razlika između grupa, potrebno je da odabremo adekvatan statistički postupak kojim ćemo to provjeriti. Prvo pitanje koje se pitamo je da li se radi o nacrtu sa paralelnim grupama ili ponovljenim mjerenjima.

6.5.1. T-test za zavisne uzorke

Kada su isti korisnici u svakoj testnoj grupi, uklonili smo glavni izvor varijabiliteta između dva skupa podataka. U takvim testovima ono što se varira je proizvod koji korisnici prvi koriste kako bi se eliminirao efekt prenošenja. Naime, ukoliko se svi korisnici prvi put susreću s proizvodom A, postoji rizik od nepravedne pristrasnosti korisnika, bilo za ili protiv proizvoda A. Prednosti su u tome što izmjerene razlike sa većom sigurnosti možemo pripisati stvarnim razlikama u korisničkom iskustvu, pri tome, na manjem uzorku.

Uz bilo kakvo poređenje, također želimo znati veličinu razlike (koja se često naziva veličinom efekta) i mogućnost greške. Kod statističkih testova možemo napraviti dva tipa grešaka:

- Greška tipa I se javlja kada u istraživanju pogrešno navedemo da je studija pronašla značajne razlike između grupa kada razlike nije bilo. Ova greška se označava sa α .
- Greška tipa II se javlja kada ustanovimo da nema razlike ili povezanosti između studijskih grupa, a u stvari ih je bilo. Ova greška se označava sa β .

Kod provođenja istraživanja unaprijed odredimo dozvoljenu marginu greške, a to se u praksi posebno odnosi na grešku tipa I, odnosno grešku da utvrdimo razlike kojih nema. Najčešće je cilj da postignemo da je greška manja od 5%. Nakon provedene studije, za svaki

test se može odrediti vrijednost greške tipa I i ova se vjerovatnoća označava kao p-vrijednost od engleskog termina za vjerovatnoću - *probability*.

Kada je P-vrijednost manja od 0.05 to znači da postoji statistički značajna razlika između mjerenja. U ovom slučaju, poredimo aritmetičke sredine dva mjerenja na istim ispitanicima. U zavisnosti od instrumenta koji smo primijenili postojanje statistički značajne razlike interpretiramo na način da se jedan proizvod karakterizira upotrebljivijim od drugog. Ipak, ostaje pitanje praktične značajnosti ovog nalaza.

Kod primjene t-testa P-vrijednost ukazuje na vjerovatnoću da postoji statistički značajna razlika između grupa, ali o razlici onda zaključujemo poređenjem srednjih vrijednosti za ove grupe i razmatramo razliku. Razlika u vrijednostima srednje vrijednosti bi mogla biti samo jedan bod na nekoj skali, što i nije od velikog praktičnog značaja. Možemo ustanoviti da neki režim vježbanja daje bolje rezultate za muškarce u odnosu na žene, i to statistički značajno sa $p=0,05$ bi značilo da će 95% muškaraca izgubiti više kilograma nego žene, ali ta razlika u izgubljenim kilogramima može biti samo 1kg. Kako veličine uzorka postaju sve veće, što je uobičajeno kod online testiranja, postaje vjerojatnije vidjeti statistički značajnu razliku kada je stvarna veličina efekta praktično možda neznčajna.

Odgovor na ovo pitanje zavisi od toga kako tumačimo najniže i najveće moguće razlike. Čak i najniža procjenjena razlika na nekoj skali može proizvod A da okarakterizira boljim od proizvoda B (kao što je slučaj na SUS standardiziranoj skali za mjerenje upotrebljivosti). Kako bismo došli do ovih zaključaka potrebno je da imamo informaciju o skali koju smo primijenili i često izmjerenim rezultatima na toj skali.

6.5.2. T-test za nezavisne uzorke

Kada dva proizvoda testiraju različite grupe korisnika, varijabilitet se javlja i između korisnika i između grupa. Svaka izmjerena razlika između grupa (npr. podataka iz upitnika, vremena zadatka) mora biti testirana da vidi da li je veća od varijanse unutar grupa.

T-test sa dva uzorka ima četiri pretpostavke:

1. Oba uzorka su reprezentativna za svoje relevantne populacije (reprezentativnost).
2. Dva uzorka nisu međusobno povezana (nezavisnost).
3. Oba uzorka su približno normalno distribuirana (normalnost).
4. Varijanse u obje grupe su približno jednake (homogenost varijansi).

Kao i kod svih statističkih postupaka, prva pretpostavka je najvažnija. P-vrijednosti, intervali pouzdanosti i zaključci su validni samo ako je uzorak korisnika reprezentativan za populaciju o kojoj izvodimo zaključke. U istraživanju upotrebljivosti to znači da pravi korisnici pokušaju izvršiti realne zadatke na pravoj verziji aplikacije.

Ispunjavanje drugog uslova obično nije problem u istraživanju korisnika jer je malo vjerovatno da će odgovori ispitanika jedne grupe utjecati na odgovore ispitanika iz druge grupe. Posljednje dvije pretpostavke, međutim, mogu biti problematične i potrebno im je posvetiti pažnju.

6.5.2.1 Normalnost distribucije

Kod t-testa na zavisnim uzorcima, t-testa na nezavisnim uzorcima i većine parametrijskih statističkih testova, postoji temeljna pretpostavka normalnosti distribucije. Konkretno, ovaj test pretpostavlja da je distribucija uzorkovanja srednje vrijednosti razlike (ne distribucija sirovih rezultata) približno normalna.

Kada distribucija aritmetičke sredine razlika nije normalna, p-vrijednosti mogu biti neodgovarajuće. Za velike uzorke (iznad 50) pretpostavka normalnosti distribucije nije problem jer je distribucija uzorkovanja srednje vrijednosti normalno raspoređena prema Centralnoj graničnoj teoremi. Ipak, uz nešto veće mjere opreza i dodatne analize, možemo reći da t-test i na manjim uzorcima daje pouzdane rezultate, čak i tamo gdje je normalnost nešto narušena.

6.5.2.2. Jednakost varijansi

Treća pretpostavka je da su varijanse (i ekvivalentno standardne devijacije) približno jednake u obje grupe. Generalno pravilo bi bilo da problem predstavlja kada je omjer između dvije standardne devijacije veći od 2 (npr. standardna devijacija od 4 u jednom uzorku i 12 u drugom je omjer 3). Kao što smo već objasnili, t-test se smatra robusnim testom, dakle ne pretjerano osjetljivim, što u praksi znači da se često i ova pretpostavka krši, posebno kada su veličine uzorka približno jednake (Sauro, 2016).

Ovdje ponovo želimo napomenuti da je presudni faktor koji ustvari određuje valjanost istraživanja, ustvari reprezentativnost uzorka. Uvijek je dobro provjeriti kakvi su podaci koje ste dobili, kako distribucija izgleda i da li imate nekih neobičnih rezultata koje trebate dodatno pogledati. Ipak, koristeći t-test većinom ćete dobiti statistički pouzdan rezultat. Pitanje valjanosti zaključaka koje donosite više ovisi o tome da li ste odabrali reprezentativan uzorak za populaciju koja vam je ciljna.

6.5.3. Interpretacija rezultata T-testa

Nakon što smo odredili tip nacрта i posljedično odgovarajuću vrstu t-testa koju ćemo koristiti, samu analizu radimo u nekom od dostupnih statističkih paketa (Excel, SPSS, Jasp i sl). Ako u timu radite sa informatičarima onda se možete osloniti i da se za analizu implementira programsko rješenje, naprimjer u jeziku R.

Rezultat koji ćemo dobiti sadrži za nas dvije bitne informacije, vrijednost t-testa i vrijednost p koja nam je bitna za interpretaciju značajnosti dobijenog rezultata. Prilikom zaključivanja o značajnosti, mi ustvari analiziramo da li je izmjerena razlika u aritmetičkim sredinama dovoljno velika na našem uzorku, da bismo bili sigurni da nije nastala slučajno. Postotak rizika od greške prilikom tog zaključivanja koji sebi dozvoljavamo je najčešće 5%. U nekim medicinskim istraživanjima se susreće i strožija dozvoljena greška - od 1%, zbog osjetljivosti predmeta proučavanja. To znači da, ukoliko je naše izmjereno $p < 0,05$ možemo zaključiti da postoji statistički značajna razlika između grupa/mjerenja. Vrijednost p koja je veća od 0,05 znači da bismo se prilikom zaključivanja da postoji statistički značajna razlika izložili riziku koji je veći od 5% da napravimo grešku i zaključimo da postoji realna razlika

nastala djelovanjem sistemskog faktora koji smo mjerili, tamo gdje je ona ustvari nastala djelovanjem slučaja.

Pregledom literature može se ustanoviti da se t-test često koristi i za mjere na bazi Likertove skale koja je ordinalna skala. Iako algoritmi odlučivanja o izboru testova tada preporučuju Mann-Whitney ili Wilcoxon test (pogledati tabelu 4.1) interesantno je pogledati rezultate studije De Wintera i Dodoua (2010) koji su uporedili greške tipa I i tipa II t-testa u odnosu na Mann-Whitney-Wilcoxon test za odgovore na Likertovoj skali sa 5 tačaka. Autori su napravili računarsku simulaciju, definisali četrnaest specifičnih populacijskih distribucija, pa iz populacija generisali preko 90 parova grupa uzoraka i podvrgnuli ih t-testu i Mann-Whitney-Wilcoxon (MWW) testu. Parovi grupa uzoraka su kreirani i da imaju balansirani broj uzoraka i sa velikim razlikama u broju uzoraka između grupa u paru. Rezultati su pokazali da su dva testa imala ekvivalentnu snagu za većinu parova. MWW je imao prednost u snazi kada je jedan od uzoraka izvučen iz iskrivljene ili multimodalne distribucije. Značajno je da je razlika u grešci tipa I kod oba testa nikada nije bila veća od 3% iznad nominalne granice od 5%, čak i kada su veličine uzoraka bile veoma nejednake. Autori su zaključili da za Likertovu skalu od 5 tačaka t-test i MWW imaju sličnu snagu, a istraživači uopšte ne moraju brinuti o pogrešnom pronalaženju razlike u grupama ako ih nema u populaciji, bez obzira koji test primijenili.

6.6. Hi-kvadrat test

Vrlo često nas u istraživanjima upotrebljivosti zanima da li je korisnik uspio izvršiti zadatak koristeći naš softver. U tom slučaju, govorimo o binarnim varijablama koje po definiciji imaju samo dvije vrijednosti: da/ne, kupljeno/nije kupljeno, zadatak uspješno završen/ zadatak nije uspješno završen i sl. Ove varijable su kodirane u vrijednosti od 1 i 0. U nekim slučajevima se i kontinuirane varijable mogu rekodirati u binarne, npr. broj korisnika kojima je trebalo manje od minute da završe zadatak, ili broj odgovora 9 ili 10 na skali od 11 tačaka. Hi-kvadrat test poredi opažene (izmjerene) frekvencije nekog odgovora npr. koliko je ispitanika zaokružilo odgovore A, B ili C, sa teorijskim (teorijske frekvencije bi bile za 30 ispitanika po 10 odgovora za svaku od alternativa). Distribuciju teorijskih frekvencija ovdje interpretiramo tj. tretiramo kao distribuciju frekvencija koju bismo dobili slučajem. Ovako nam Hi-kvadrat daje odgovor na pitanje da li se naša distribucija frekvencija stastistički značajno razlikuje od teorijski očekivane.

Kao i kada smo govorili o poređenju dvije aritmetičke sredine, prvi korak u odlučivanju koji postupak koristiti je odrediti da li se radi o zavisnom (jedna grupa ispitanika u više tačaka mjerenja) ili nezavisnom (dvije i više grupa u različitim situacijama) uzorku.

Naredni korak koji trebamo razmotriti je veličina uzorka. Hi-kvadrat test daje pouzdanije rezultate na većim uzorcima. Pravilo koje je se navodi u statističkim udžbenicima jeste da veličina uzorka za Hi-kvadrat test treba biti minimalno tolika da svaka od mogućih alternativa ima teorijsku frekvenciju makar 5. Dakle, ukoliko govorimo o tri moguće alternative odgovora (A, B ili C) teorijske frekvencije trebaju biti 5, ukupno 15 ispitanika.

Nakon što smo identificirali da imamo dovoljno velik uzorak, i odredili koja vrsta analize nam treba, koristeći neki od statističkih softvera (MS Excel, SPSS) izračunamo vrijednost Hi-kvadrata i pripadajuću p vrijednost.

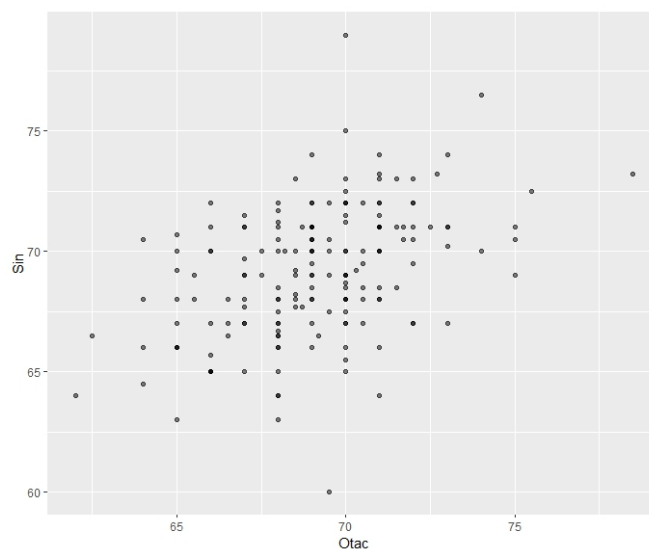
Podsjećamo da i ovaj test ukazuje na postojanje statistički značajne razlike u mjerenjima između grupa koje su podvrgnute testu upotrebljivosti odnosno korisničkog iskustva. Nakon što smo ustanovili razliku, možemo pristupiti poređenju metrika, odnosno stavova prema upotrebljivosti određenog softvera između grupa.

Postoji dosta diskusije u literaturi vezano za upotrebu Hi-kvadrat testa za mjere ordinalne skale kao što je npr. Likertova skala slaganja. Iako to nije uputno i ne slaže se sa preporukama algoritama odlučivanja o dizajniranju studija, jednostavnost primjene ovog testa motivira mnoge istraživače da ga koriste. U slučaju primjene Hi-kvadrat testa za mjere Likertove skale preporuka je analizirati odgovore ali grupisane u tri kategorije: slaganje, neslaganje i neutralno. Ovi odgovori se mogu transformisati i u binarni tip varijable tako da neutralne odgovore pridružimo onoj grupi odgovora koja nam nije poželjna. Važno je istaći da se kod obrade odgovora treba voditi računa o načinu postavljanja pitanja, i za pitanja koja su sa negativnim pristupom odgovori se komplementiraju prije pridruživanja određenoj grupi. Primjer razlike u pristupu su pitanja „Komande za kretanje u VR okruženju su me zbunjivale.“ i „Snalaženje u VR okruženju je bilo jednostavno“.

Sada ne možemo računati srednju vrijednost, već se koristimo metrikama na bazi frekvencija kao što su ukupan procenat odgovora slaganja (eng. *Percent Agree*), procenat odgovora maksimalnog slaganja (eng. *Top Agree*), ili Net Promotor Score, koji se računa kao razlika ukupnog procenta odgovora slaganja i ukupnog procenta odgovora neslaganja.

6.7. Korelacija

Osim frekvencija, korelacija je još jedna mjera koja se često koristi u istraživanjima za procjenu potencijalnih veza između varijabli. Naime, kada procjenjujemo korelaciju između dvije varijable, mjerimo koliko su te varijable povezane jedna s drugom.



Slika 6.13. Dijagram raspršenja kojim se uvezuju visina očeva i sinova

Na primjer, ako radimo na web stranici za e-trgovinu, možda bismo željeli znati postoji li korelacija između toga koliko dugo neko ostaje na stranici i kolika je vjerovatnoća da će nešto kupiti na toj stranici. Ili se možemo zapitati: postoji li korelacija između toga koliko web stranica klijent posjeti i koliko na kraju potroši?

Ključni alat za vizualizaciju korelacije između dvije varijable je dijagram raspršenja. Dijagram raspršenja je jednostavna vizualizacija koja upoređuje dvije numeričke varijable koristeći tačku za predstavljanje svakog pojedinačnog podataka. Ove tačke podataka se određuju tretiranjem odgovarajućih vrijednosti dvije varijable kao koordinata na grafu, npr. korisnik koji je proveo 30 minuta pregledajući web stranicu i napravio dvije kupovine ima tačku u koordinatnom sistemu - (2,30). Svakog ispitanika na isti način možemo pozicionirati u koordinatnom sistemu i ustvari imati kompletan skup podataka grafički predstavljen. Na slici 6.12. je prikazan dijagram raspršenja za visine očeva i sinova iz Galtonove studije analize visina roditelja i djece (Galton, 1886).

6.7.1. Koeficijent korelacije

Koeficijentom korelacije izražavamo jačinu i smjer korelacije. Smjer korelacije se izražava predznakom, + ili -. Pozitivna korelacija znači da rast jedne varijable prati rast druge i obrnuto, opadanje vrijednosti jedne dovodi do opadanja vrijednosti druge varijable. Negativna korelacija također govori o sukladnom variranju varijabli ali na način da rast vrijednosti na jednoj varijabli dovodi do srazmjernog opadanja vrijednosti druge varijable, i obrnuto.

Osim razlikovanja smjera korelacije, tj. pozitivne i negativne korelacije, korelaciju mjerimo i po jačini povezanosti. Dakle, imamo način da precizno kažemo kolika je korelacija između dvije varijable. To možemo učiniti tako što ćemo izračunati koeficijent korelacije za naš skup podataka. Koeficijent korelacije je numerička vrijednost, u rasponu između -1 i 1, gdje -1 predstavlja maksimalnu negativnu korelaciju, a 1 maksimalnu pozitivnu korelacije. Naravno, postoji 0 tačka između, što ne predstavlja nikakvu korelaciju u podacima. Evo tipične skale koeficijenta korelacije koja se koristi za opisivanje količine korelacije naznačene izračunatim koeficijentom:

Tabela 6.2. Skala korelacije

Raspon koeficijenta korelacije	Zaključak
0 - 0.3	ne postoji povezanost
0.3 - 0.5	niska poveznost
0.5 - 0.7	srednja povezanost
0.7 - 1	visoka povezanost

6.8. Faktorska analiza

Faktorska analiza (FA) je statistički postupak kojim se identificira koje varijable formiraju koherentne grupe relativno nezavisne jedne od drugih. Drugačije rečeno, FA je tehnika koja omogućava svodenje većeg seta manifestnih varijabli na manji set latentnih varijabli koje su u osnovi testnih rezultata. Osnovna pretpostavka od koje polazimo u

faktorskoj analizi je ako između dvije varijable postoji visoka korelacija, onda postoji zajednički faktor koji djeluje na uradak i u jednoj u drugoj varijabli. Cilj faktorske analize je pronaći zajedničke faktore koji su u osnovi manifestnih varijabli. Faktorskom analizom nastojimo odgovoriti na tri ranije spomenuta pitanja koliko zajedničkih faktora učestvuje u formiranju testnih rezultata, koje je psihološko značenje tih faktora i kakav je odnos među faktorima? Na prvo pitanje odgovaramo kroz ekstrakcije faktora iz manifestnih varijabli i pregledom objašnjene varijance, na drugo pitanje odgovaramo pregledom povezanosti faktora s manifestnim varijablama, a na treće pitanje odgovaramo rotacijom faktora. Faktorsku analizu dijelimo na eksploratornu i konfirmatornu. U eksploratornoj FA osnovni cilj je utvrđivanje zajedničkih latentnih faktora koji leže u osnovi korelacija što većeg broja manifestnih varijabli. Njen cilj je deskriptivan, tj. faktorski opis interkorelacija što većeg broja opaženih varijabli. Konfirmatorna FA polazi od unaprijed zadanog modela strukture rezultata i broja latentnih faktora, a glavni cilj je potvrđivanje modela.

Ilustrativni primjer problema sa kojim nam faktorska analiza pomaže su tekstualni zadaci na provjerama znanja iz matematike. Vrlo često, od djece se očekuje da zadatak koji je dat tekstualno, koristeći logiku, postavbe i kroz matematičke simbole predstave, pa dalje dođu do rješenja. S obzirom na to da je zadatak predmetom provjere znanja iz matematike, pretpostavka je da se njime mjeri matematičko znanje. Ipak, da li takav zadatak mjeri samo znanje iz matematike? Zamislimo dijete koje je iz inostranstva i jezik koji koristi u školi nije njegov/njen maternji. Da li to dijete ima dodatnu teškoću u razumijevanju toga šta se od njega traži u odnosu na dijete koje ide u školu na maternjem jeziku. Moguće je da neku riječ ne razumije i općenito moguće je da ima teškoće u logičkom slaganju pojmova na stranom jeziku. Sličan problem se može javiti i kod djece koja imaju neke jezičke teškoće. Rezultat na testovima matematičke logike, koji sadrže ovakve zadatke može biti nerealno niži jer, ako razmislimo, ovakvi zadaci ne mjere samo matematičku sposobnost/znanje nego i jezičku kompetentnost. Ovdje dolazimo do tog pitanja na koje nam faktorska analiza daje odgovor. Da li mi stvarno mjerimo ono što mislimo da mjerimo? Ovakvi zadaci, u tablici rezultata eksploratorne faktorske analize se pokazuju kao faktorski nečisti. Komunalitet, tj varijansa odgovora na svakoj pojedinačnoj čestici u totalu iznosi 1. Ta totalna varijansa se sastoji od dijela rezultata za koji je odgovoran faktor čije smo djelovanje pretpostavili (u našem primjeru je to faktor znanja iz matematike), dijela rezultata koji je odredio neki drugi faktor (npr. verbalni) i dijela koji otpada na grešku. Mi naravno želimo da je najveći dio totalne varijanse (dakle vrijednost najbliža 1) ona koju mi i mislimo da mjerimo. Koristeći postupak eksploratorne faktorske analize mi totalnu varijansu dobijemo raščlanjenu upravo na ove dijelove i možemo na osnovu toga zaključiti koliko nam je pretpostavka o mjerenju ispravna.

Konfirmatorna faktorska analiza je postupak u kojem mi definiramo unaprijed stavke za koje očekujemo da su zasićeni istim faktorom i koristeći postupak strukturalnog modeliranja testiramo tu hipotezu, tj kako se taj cijeli postavljeni model uklapa sa podacima.

7. Modeli prihvatanja tehnologije

U jednoj epizodi popularne serije Teorija velikog praska, Sheldon, jedan od glavnih likova, pokušava da odluči koju konzolu za igranje da kupi. Cijelu epizodu izvor komičnih situacija je njegovo nabranje različitih tehničkih karakteristika trenutno dostupnih modela na tržištu i ponderiranje istih po važnosti. Negdje na kraju epizode obeshrabren svojom nemogućnošću da odabere, navodi svoje prethodne pogrešne procjene glede tehnologije koja će postati dominantna na tržištu (npr. Investiranje u Zune mp3 player ili HD DVD).

Dilema koju Sheldon ima u ovoj epizodi je ustvari slika u ogledalu zadatka koji imaju stručnjaci iz područja informacione tehnologije (IT) prilikom dizajniranja bilo kog proizvoda za tržište. Kako što smo u više navrata u prethodnim poglavljima navodili, i IT proizvodi i usluge, kao i sve ostale, na kraju, svoju uspješnost mjere time koliko ih korisnici prihvate i koriste. Kako bismo što bolje dizajnirali proizvod, na način da bude što upotrebljiviji, koristimo različite, već opisane tehnike i postupke mjerenja upotrebljivosti. Međutim, problem upotrebljivosti je potrebno sagledati i sa još jednog aspekta, a taj aspekt je vezan za općenito nivo prihvatanja korištenja tehnologije za izvršavanje nekog zadatka od strane korisnika.

Davis i Granić (2024) navode kako je 1980tih godina prošlog stoljeća, odbijanje uvođenja novih sistema organizacije poslova zasnovanih na IT, čak i u situacijama kada je tehnologija značajno olakšavala i ubrzavala obavljanje aktivnosti, navelo IT stručnjake na zaključak kako je nemoguće predvidjeti uspjeh nekog proizvoda i tehnologije, tj da nije moguće predvidjeti ljudsko prihvatanje određene tehnologije u svoju svakodnevicu. Mogućnost razvoja modela koji bi mapirao faktore koji utječu na prihvatanje određene tehnologije od strane korisnika je postalo i ekonomsko pitanje. Ukoliko su pojedini faktori vezani za odlučivanje o korištenju određene tehnologije poznati, moguće je sve resurse usmjeriti na razvoj tehnologije koja ima najveću vjerovatnoću uspjeha. Ulaganje resursa u tehnologiju koju korisnik neće prihvatiti, u konačnici, predstavlja bacanje novca, što nije dobrodošla pojava u biznisu.

Generalno, mogućnost predviđanja ljudskog ponašanja u bilo kom segmentu pa tako i u segmentu prihvatanja tehnologije, neraskidivo je vezana za napredovanje i razvoj istraživačkih postupaka kojim mjerimo upravo ljudsko ponašanje.

U ovom poglavlju ćemo predstaviti dva trenutno dominantna modela prihvatanja tehnologije, TAM - *Technology Acceptance Model* (Davis 1986, 1989) i UTAUT - *Unified Theory of Acceptance and Use of Technology* (Venkatesh i sur., 2003). Navedeni modeli pretpostavljaju i operacionaliziraju individualne faktore koji doprinose prihvatanju tehnologije, a koji su vezani za karakteristike, motivaciju i potrebe i korisnika.

7.1. Model prihvatanja tehnologije - TAM

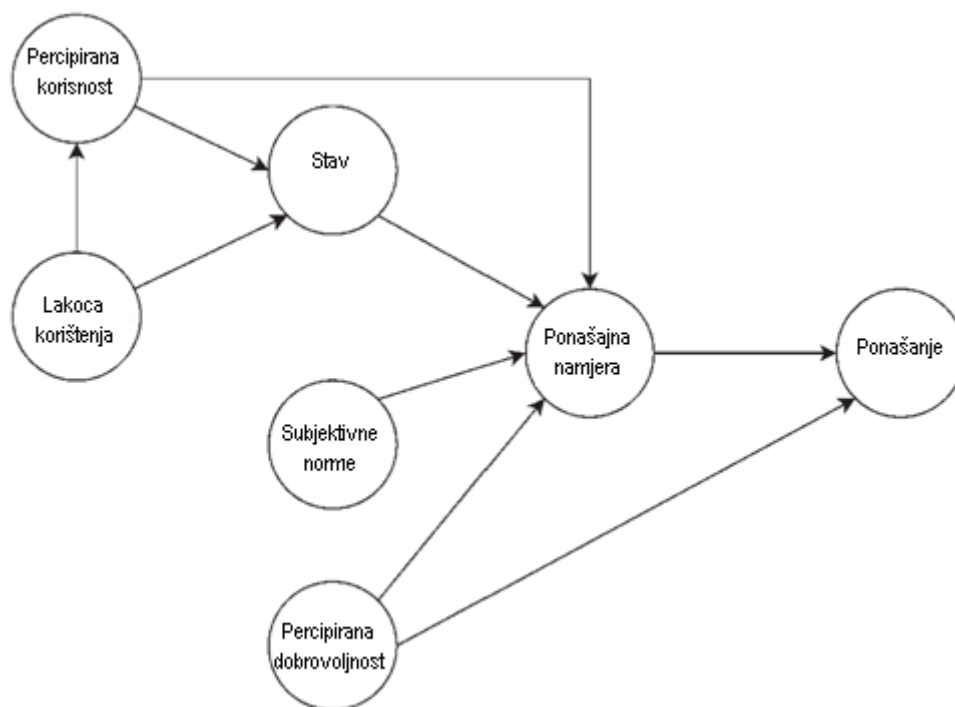
Davis (2024) navodi kako su ciljevi TAM modela: identifikacija ključnih determinanti prihvatanja tehnologije od strane korisnika (percipirana korisnost i percipirana jednostavnost korištenja) i razvijanje valjanih i pouzdanih mjera za percipiranu korisnost i percipiranu lakoću korištenja.

Validirane mjere percipirane korisnosti i jednostavnosti korištenja uspješno predviđaju i objašnjavaju kako objektivne razlike u dizajnu proizvoda utiču na vjerovatnoću prihvatanja od strane korisnika. Uočena korisnost je primarna determinanta prihvatanja korisnika, a percipirana jednostavnost korištenja je moćna sekundarna determinanta koja utiče na prihvatanje korisnika i direktno i indirektno kroz svoj utjecaj nauočenu korisnost. Primat percipirane korisnosti nad lakoćom upotrebe je logičan. Korisnici će tolerisati neke poteškoće u korišćenju sistema ako on omogućava izvršavanje aktivnosti koje se doživljavaju vrlo korisnim. S druge strane, ne postoji lakoća upotrebe koja može nadoknaditi nedostatak percipirane korisnosti. Ovaj nalaz opravdava testiranje prihvatljivosti tehnologije ranije u njenom razvoju. Naime, s obzirom na to da je percipirana korisnost jači prediktor prihvatanja od lakoće upotrebe, ne moramo čekati da proizvod u kompletu bude završen da bismo testirali kako ga korisnici prihvataju. Percipiranu korisnost je moguće mjeriti i testirati i na proizvodu kojem nisu neki operativni dijelovi do kraja razrađeni. Time se opet otvaraju mogućnosti uštede i intervencija na proizvodu dok je u fazi izrade, kako bi se što bolje prilagodio potrebama tržišta.

Model prihvatanja tehnologije (TAM), koji je prvobitno predložio Davis u svojoj doktorskoj disertaciji (1986.) prvenstveno se fokusira na uočenu korisnost i percipiranu lakoću upotrebe tehnologije kao ključne determinante usvajanja tehnologije. Međutim, TAM ne uključuje utjecaj društvenih uvjerenja i osjećaja kontrole na ponašanje, faktora koji imaju empirijski dokazano značajan uticaj na ponašanje prilikom upotrebe tehnologije (ustvari, koji generalno imaju utjecaj na ljudsko ponašanje u mnoštvu aktivnosti pa tako i u upotrebi tehnologije).

Kako bi se prevažili uočeni nedostaci TAM modela, razvijen je tkz. A-TAM (*Augmented Technology Acceptance Model*), koji inkorporira faktore bihevioralne kontrole i subjektivnih normi koji potpisuju Taylor i Todd (1995, prema Davis i Granić, 2024). Ovaj model je prikazan na slici 7.1.

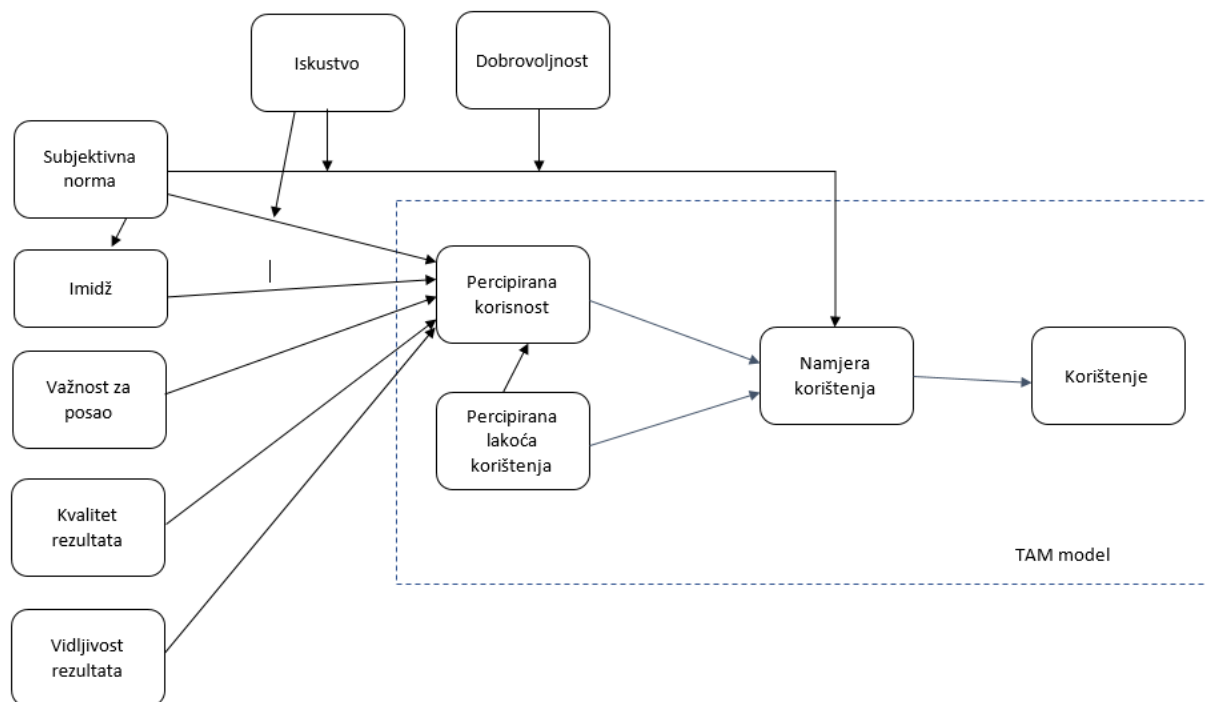
U svjetlu robusnih istaživačkih nalaza koji govore u prilog velikom značaju percipirane korisnosti prilikom upotrebe tehnologije, Vekantesh i Davis (2000) razvijaju dodatno TAM model, koji će nazvati TAM-2. Naslanjajući se na osnovu TAM-a, TAM2 uvodi dodatne teorijske konstrukte koji imaju za cilj da objasne determinante percipirane korisnosti i upotrebne namjere kroz socijalne i kognitivne utjecaje.



Slika 7.1. Model A-TAM

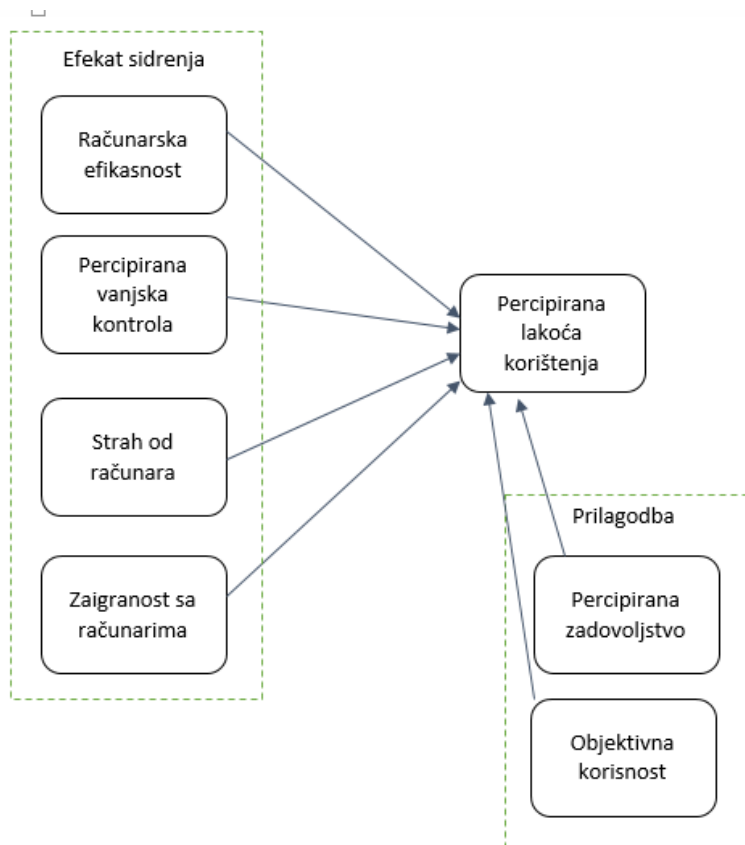
TAM2 obuhvata uticaj tri međusobno povezana društvena faktora koji utječu na pojedinca kada donose odluku da prihvate ili odbace novu tehnologiju: subjektivne norme, dobrovoljnost i imidž. Pored razjašnjavanja procesa utjecaja društvenih faktora koji oblikuju percipiranu korisnost i upotrebnu namjeru, autori postavljaju četiri kognitivne instrumentalne determinante koje dodatno operacionaliziraju uočenu korisnost: relevantnost posla, kvalitet rezultata, vidljivost rezultata i percipirana jednostavnost upotrebe. U TAM-2 modelu se uključuju i iskustvo i dobrovoljnost kao moderirajući faktori. Značajni nalazi iz istraživanja u svrhu validacije TAM- 2 modela uključuju moderaciju efekta prethodnog iskustva na subjektivno percipirane norme, kao i značajan uticaj iskustva i percipirane dobrovoljnosti na subjektivni odnos normi i namjera. Slika 7.2. predstavlja TAM-2 model.

Analizom TAM-2 modela ostalo je otvoreno pitanje da li determinante percipirane korisnosti mogu uticati na percipiranu lakoću upotrebe, ili obrnuto. Kako bi se dodatno objasnio fenomen prihvatanja tehnologije, razvijen je i TAM- 3 model koji nudi “sveobuhvatnu nomološku mrežu” koja obuhvata determinante usvajanja i korištenja tehnologije na individualnom nivou.



Slika 7.2. Model TAM-2

Važno je da TAM-3 uvodi tri ključna odnosa koja do sada nisu bila empirijski istražena u prethodnim modelima. Autori pretpostavljaju moderirajući utjecaj faktora prethodnog iskustva u odnosu između percipirane lakoće upotrebe i percipirane korisnosti, anksioznosti od računara i percipirana lakoće upotrebe, kao i percipirane lakoća upotrebe i ponašajne namjere (Davis i Granić, 2024). Model TAM-3 stavlja poseban naglasak na značajnu ulogu procesa povezanih sa percipiranom korisnošću i percipiranom lakoćom upotrebe. U okviru TAM-3 modela pretpostavlja se da determinante koje utiču na percipiranu korisnost neće uticati na percipiranu lakoću upotrebe i obrnuto. Nadalje, TAM-3 pretpostavlja da kako se iskustvo sa tehnologijom akumulira, uticaj percipirane lakoće upotrebe na namjeru ponašanja se smanjuje, dok uticaj percipirane lakoće upotrebe na percipiranu korisnost raste. Ovo naglašava longitudinalni utjecaj percipirane lakoće upotrebe na odgovor korisnika na informacijsku tehnologiju, čak i kada se radi o korisnicima koji posjeduju znatno praktično iskustvo. Slika 7.3. prikazuje TAM-3 model.



Slika 7.3. Proširenje uticaja na Percipiranu lakoću korištenja kod modela TAM-3

Model prihvatanja tehnologije (TAM) našao je široku primjenu zahvaljujući rasprostranjenom konsenzusu vezanom za njegovu valjanost i pouzdanost u predviđanju korisničkog prihvatanja niza proizvoda i usluga informacione tehnologije. Upotreba TAM modela je bila predmetom istraživanja u okviru širokog spektra tehnoloških rješenja, interaktivnih sistema, okruženja, alata, aplikacija, usluga i uređaja.

7.2. Objedinjeni teorijski model prihvaćanja i korištenja tehnologije - UTAUT

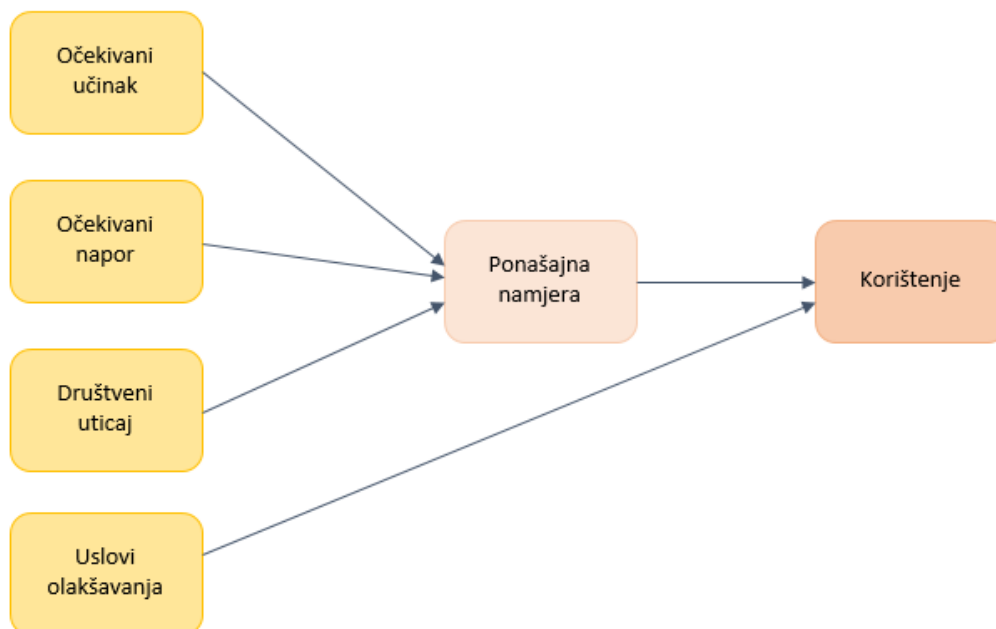
UTAUT teoriju su razvili Venkatesh i sur. (2003) u pokušaju da prevaziđu nedostatke prethodnih teorija i modela. UTAUT teorija inkorporira osam najutjecajnijih modela i teorija prihvatanja tehnologije koje su joj prethodili. Venkatesh i sur. (2003) su formulisali i novi model koji su nazvali također UTAUT, u okviru kojeg su definirana tri konstrukta koji pokazuju ključnu namjeru za korištenje određene informacione tehnologije: očekivani napor, očekivani učinak i društveni uticaj. Ova tri konstrukta obuhvataju najrobusnije konstrukcije osam prethodnih teorija i modela na kojima se UTAUT zasniva.

Očekivani učinak se donosi na očekivanje korisnika da će im korištenje datog sistema pomoći da povećaju vlastiti učinak poslu. Ovaj konstrukt ima korjene u svim verzijama TAM modela, postoji jasna veza sa konceptom ekstrinzične motivacija iz motivacionih modela kao koncepta očekivanja/ishoda iz socijalno-kognitivne teorije. Venkatesh i sur (2003) navode kako je ovaj konstrukt najjači prediktor kojim se predviđa upotreba tj. prihvatanje određene tehnologije. Logično je pretpostaviti da postoji moderacijski efekat dobi i spola na odnos

između varijabli očekivani učinak i namjera da se tehnologija koristi. Npr. muškarci su češće motiviraniji za formalno završavanje zadatka, pa se kod njih očekuje jače djelovanje faktora očekivanog učinka. Također, mlađi uposlenici su skloniji da daju veći značaj vanjskim, ekstrinzičnim nagradama. *Očekivani napor* je nivo složenosti koji je vezan za primjenu određenog sistema. *Društveni uticaj* je nivo na kojem osoba percipira da relevantni pojedinci vjeruju da se predmetna tehnologija treba koristiti. Relevantni pojedinci su važni u poslovnom okruženju, to su nadređeni, ali i nama jednaki, jer nas i oni procjenjuju kao profesionalce i stručnjake. Konstrukt društvenog uticaja je važan i za aplikacije koje se koriste u slobodno vrijeme, posebno kada se koriste u komunikaciji kao što su društvene mreže.

Venkatesh i sur. (2003) su definirali efekat *olakšavajućih uslova* za primjenu određene tehnologije. Ovaj faktor se odnosi na uvjerenje pojedinca da postoji organizacijska i tehnička podrška za korištenje nove tehnologije. Ova uvjerenja su prvenstveno određena iskustvom i godinama pojedinca. Niska vrijednost ovog koncepta može ukazati na nedostupnost korištenja određene tehnologije zbog cijene, problema nabavke, prepreka u korištenju. *Ponašajna namjera* je također identificirana kao značajan prediktor stvarnog ponašanja pojedinca.

Cilj UTAUT modela je da integriše fragmentirane teorije i istraživanja koja se bave ponašanjem pojedinca kada govorimo o prihvatanju informacione tehnologije, u jedinstveni teorijski model koji obuhvata bitne elemente ranije uspostavljenih modela. Slika 7.4. grafički predstavlja UTAUT model.



Slika 7.4. Model UTAUT

Alshammari i Rosli (2020) u pregledu postojećih modela prihvatanja tehnologije dolaze do zaključka da još uvijek nije napravljen stvarno sveobuhvatan model, te da svakom od trenutnih modela nedostaju određeni faktori. Ipak, zaključak koji se nameće je da upravo

UTAUT model dolazi najbliže ka obuhvatnom modelu, te da dosljedno u istraživanjima objašnjava najveći dio varijanse ponašanja korisnika (oko 70%).

Venkatesh i sur. (2003) uključili su ključne elemente TAM modela u model UTAUT i to su očekivani učinak, očekivani napor, društveni uticaj i uslovi olakšavanja korištenja, kao i četiri ključna moderatora kao što su spol, godine, dobrovoljnost i iskustvo. Lai (2017) navodi da bi objašnjenje eksplanatorne snage UTAUT modela moglo biti u njegovoj robusnoj, moćnoj statističkoj strukturi koja omogućava objašnjavanje velikog procenta varijanse sa relativno malim brojem varijabli, ali model ne ispituje direktne efekte koji mogu otkriti nove odnose kao i važne faktore iz studije koji su izostavljeni u svrhu racionalizacije broja prediktora.

8. Teorija kognitivnog opterećenja i upotrebljivost

Teorija kognitivnog opterećenja nastala je osamdesetih godina prošlog stoljeća. Riječ je o kognitivnoj teoriji koja je primarno nastala sa svrhom dizajniranja situacije podučavanja kako bi se olakšalo učenje.

Razvoj IT je doveo do toga da različite digitalne alate svakodnevno koristimo kao didaktička sredstva. Razvoj obrazovnog softvera je omogućio ne samo hibridno učenje, nego i potpuno digitaliziranu nastavu i poduku. Dizajniraju se igrice koje su edukativnog karaktera i imaju za cilj usvajanje i razvoj različitih vještina kod djece. Ove igre koje se dizajniraju za učenike različitih uzrasta prepoznaju se kao posebna vrsta softvera pod nazivom „ozbiljne igre“. Pojedinci na tržištu rada vrlo često moraju usvojiti korištenje novih digitalnih alata kako bi uspješno obavljali posao ili razumjeli nove koncepte, i često se obučavaju upravo korištenjem ozbiljnih igara. Ovakve promjene u našoj svakodnevnici, koje je donio razvoj IT zahtijevaju dublje razumijevanje načina na koji ljudi uče, kako bi im se olakšalo da drže korak sa stalnomijenjajućom stvarnosti. Dakle, da bi korisnik prihvatio određenu tehnologiju, da bi percipirani napor da se savlada korištenje (koji je identificiran kao značajan faktor za prihvatanje tehnologije) bio što manji, potrebno je razumijeti kako je postavljena čovjekova kognitivna arhitektura i u skladu s tim dizajnirati proizvode.

Autor Teorije kognitivnog opterećenja (TKO), John Sweller, argumentira da je intervencijom prilikom planiranja situacije podučavanja moguće značajno poboljšati efikasnost podučavanja. Kao što je navedeno, sama teorija je nastala u okviru kognitivne psihologije i psihologije učenja i podučavanja, pa će na takav način i biti objašnjena. Kasnije u poglavlju ćemo napraviti jasne poveznice sa konceptom upotrebljivosti posebno obrazovnog, ali i drugih vrsta softvera.

Prema Sweller, Ayres i Kalyga (2011), instrukcija prilikom podučavanja treba da bude dizajnirana na način da je u skladu sa kognitivnom arhitekturom čovjeka. Samo u tom slučaju će kognitivni kapacitet biti optimalno iskorišten i proces podučavanja će biti efikasan.

Značajan korpus istraživanja dizajniranih sa ciljem provjere pretpostavki teorije doveo je do razvijanja empirijski provjerenih smjernica za dizajn instrukcije u podučavanju. Stoga je teorija kognitivnog opterećenja od samog svog pojavljivanja postala utjecajna teorija koja se bavi dizajnom instrukcije, a u svojim narednim fazama razvoja i dizajnom materijala za učenje te multimedijalnog učenja.

Jedna od osnovnih pretpostavki teorije kognitivnog opterećenja (TKO) je da dizajn instrukcije treba biti u skladu s kognitivnom arhitekturom čovjeka kako bi se poboljšalo učenje. Van Merriënboer i Sweller, J. (2010) navode kako je TKO od samog početka nudila neka rješenja koja su za nastavno osoblje bila kontraintuitivna. Na primjer, uputa da se početnicima ponudi puno urađenih primjera. Ova uputa je u nesuglasju sa vladajućim uvjerenjem da se najbolje uči kroz samostalno rješavanje problema. Od svog nastanka, TKO se razvijala tako da danas nudi različite empirijski provjerene smjernice za dizajn instrukcije,

kako bi se optimizirala usmjerenost kognitivnih resursa i time olakšalo učenje u različitim situacijama.

S druge strane, upotreba tehnologije u nastavi je odavno stvarnost a ne neka hipotetička mogućnost. U početku je tehnologija imala isključivo ulogu medija preko kojeg se određeni sadržaj učenicima prezentira radi lakšeg razumijevanja. Danas, imamo razvijene sofisticirane aplikacije za učenje kao i online kurseve gdje polaznici mogu steći napredna znanja iz nekog područja i na kraju kursa dobiti certifikate koji su nerijetko prepoznati na tržištu rada kao relevantni. Značaj upotrebe, ali i unapređenja tehnologije koja se koristi u obrazovanju je posebno aktualiziran nakon pandemije COVID-a gdje je u jednom trenutku kompletan nastavni proces prebačen u format učenja na daljinu, te se učenje i podučavanje dešavalo isključivo uz pomoć tehnologije. Ovakve okolnosti su značajno aktualizirale pitanje upotrebe obrazovnih tehnologija i potrebu za njihovim unapređenjem.

S tim u vezi, značajna je uloga istraživanja iz područja upotrebljivosti softvera koja se rade u okviru procesa razvoja svakog softvera. Lazar (2006) upotrebljivost definira kao stepen u kojem proizvod može da se koristi od strane određenog korisnika kako bi se efikasno, efektivno i sa osjećajem zadovoljstva postigao određeni cilj relevantan za kontekst upotrebe.

Dakle, svaki softver da bi ispunio svoju funkciju mora biti upotrebljiv, tj. prilagođen korisniku na način da mu omogući efikasno i efektivno izvršavanje zadataka. Da bi to bio u mogućnosti, dizajn softvera mora biti u skladu sa kognitivnom arhitekturom i načinima kako ljudi procesiraju informacije. S obzirom na to da se Teorija kognitivnog opterećenja upravo bavi definiranjem kognitivne arhitekture i mapiranjem načina obrade informacija kako bi se maksimiziralo učenje, logičan zaključak je da se prilikom dizajniranja obrazovnog softvera treba inkorporirati principe i saznanja do kojih se došlo u okviru TKO.

8.1. Teorija kognitivnog opterećenja

Plass, Moreno i Brünken (2010) navode da je cilj TKO objašnjavanje psiholoških i bihevioralnih fenomena do kojih dolazi kao rezultat instrukcije. Psihološke teorije se bave, generalno, izučavanjem mogućih odnosa psiholoških konstrukata između sebe ili između psiholoških konstrukata i opaženih fenomena od praktičnog značaja. Konstrukti u centru TKO su kognitivno opterećenje i učenje. TKO nastoji objasniti utjecaj dizajna instrukcije na ova dva konstrukta.

Konstrukt kognitivnog opterećenja bio je poznat i prije TKO. Sličan konstrukt, mentalno opterećenje, već je bilo predmet istraživanja. Na primjer, Plass, Moreno i Brünken (2010) objašnjavaju da je mentalno opterećenje definirano kao razlika između zahtjeva zadatka i kapaciteta osobe da ispuni te zahtjeve. Odnos između mentalnog opterećenja i postignuća već je istraživao u različitim područjima, kao što su kognitivna ergonomija, informacione tehnologije (npr. u domenu dizajna i iskoristivosti softvera). Također, istraživanja vezana za konstrukt težine zadatka koji se gotovo uvijek razmatra u odnosu na mentalno opterećenje su dovela do saznanja o procjeni težine posla (tj. radnih zadataka) u različitim zanimanjima. Utjecaj ovih saznanja na TKO je značajan. Na primjer, ideja o konstrukciji prve skale subjektivnog mjerenja kognitivnog opterećenja, koju su ponudili Paas

i van Merriënboer (1994), bila je potaknuta prethodno dizajniranim skalama za procjenu percipirane težine zadatka te sličnim kognitivnim testovima.

Mentalno opterećenje nije povezano samo sa količinom posla koji treba uraditi da se uspješno završi zadatak. Drugi psihološki faktori, kao što su očekivanja, uloženi trud i percipirana kompetentnost za završavanje zadatka su važni u predviđanju mentalnog opterećenja. Na primjer, ukoliko je za završetak nekog zadatka potrebno uložiti puno truda, različito opterećenje koje će se javiti kod različitih osoba vezano je za njihovu motivaciju za ulaganjem truda u taj zadatak. Motivacija je pod utjecajem slike o sebi i procjene koliko je zadatak važan za postizanje vlastitih ciljeva.

8.1.1. Organizacija pamćenja

Prema Sweller, Aryes i Kalyuga (2011), kognitivna arhitektura se odnosi na način na koji se kognitivne komponente organiziraju u ljudskom kognitivnom sistemu.

U okviru TKO navodi se kako su osnovne karakteristike ljudske memorije: snaga i trajnost, kapacitet i brzina pristupa. S obzirom na ove karakteristike, pamćenje je podjeljeno na kratkoročno tj. radno pamćenje i dugoročno pamćenje. Dugoročno pamćenje karakterizira velika snaga, sadrži dobro naučeno znanje i pretpostavlja se da ima neograničen kapacitet, iako pristup nekim informacijama može biti spor. Snaga dugoročnog pamćenja i brzina pristupa se povećavaju sa vježbom.

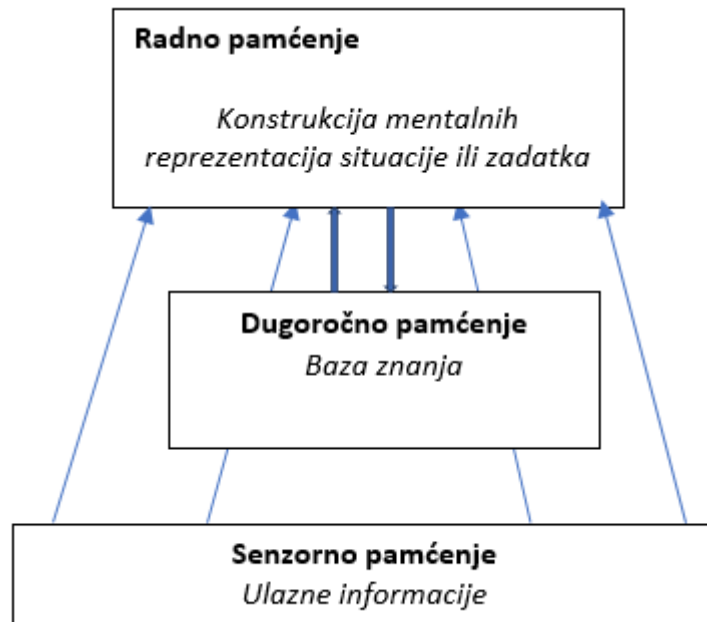
Radno pamćenje uključuje informacije koje su upravo dospjele u senzorne registre ili su pozvane iz dugoročne memorije. Trajanje radne memorije je nekoliko sekundi, i informacije u radnoj memoriji se brzo obrađuju. Broj informacija koje se mogu istovremeno obrađivati u radnom pamćenju je oko sedam za većinu ljudi (Miller, 1956), premda nailazimo i na podatak o četiri informacije (npr. Cowan, 2010).

Kako bi se adekvatno objasnila kognitivna arhitektura, potrebno je dodati i centralni procesor koji određuje koje će informacije iz senzornih registara ili dugoročnog pamćenja biti pozvane u radnu memoriju i obrađene. Jedan od mogućih načina procesiranja informacija izgleda ovako: kada je stimulus iz okoline registrovan u senzornom registru, percipiran ili prepoznat s obzirom na prethodno pohranjenje informacije u dugoročnom pamćenju, u radnu memoriju se pozivaju već dostupne relevantne informacije o stimulusu. U radnoj memoriji se uočene karakteristike slažu u mentalnu reprezentaciju situacije ili zadatka. Sve ove informacije će brzo nestati iz radne memorije, ukoliko se pažnja sa njih pomjeri ili dođe do preopterećenja radnog pamćenja sa prevelikim brojem elemenata. Slika 8.1 shematski prikazuje put informacija kroz elemente kognitivne arhitekture.

Individualne razlike u kapacitetu radne memorije koje za određene zadatke dovode do razlika u izvedbi su još uvijek predmet istraživanja. Ove razlike mogu biti pod jakim utjecajem znanja u dugoročnoj memoriji. Kapacitet radne memorije ustvari reflektira pojedinčevo poznavanje određenog područja, a ovo prethodno znanje neupitno utječe na obradu i pohranu informacija.

Mjere kapaciteta radne memorije u jednoj domeni se mogu koristiti kao prediktori uspješnosti izvedbe u bliskim područjima djelovanja, a ne kao opća mjera kapaciteta radne

memorije. Praktično je nemoguće eliminirati utjecaj prethodnog znanja u situacijama kada se testira uspješnost na smislenim zadacima radnog pamćenja. Pristupi sa fokusom na vezu između sadržaja, procesa radnog pamćenja i dugoročnog pamćenja pokazuju se kao produktivniji u odnosu na ostale pristupe (Kalyuga, 2009).



Slika 8.1. Osnovna kognitivna arhitektura (Kalyuga, 2009)

Prethodno znanje iz predmetnog područja omogućava ekspertima da kodiraju i zadrže veliku količinu informacija u dugoročnoj memoriji. Operacije pohrane i pozivanja informacija iz dugoročne memorije vježbom postaju brže, što za rezultat ima superiorni uradak koji eksperti pokazuju prilikom rješavanja problema i dozivanja informacija iz dugoročne memorije. Npr., eksperti i mnemoteknikama mogu povećati broj znamenki koje pamte na puno veći broj od prosječnih sedam. Oni koriste već poznate informacije iz dugoročne memorije za kodiranje novih informacija u lako dostupnom obliku.

8.2. Vrste kognitivnog opterećenja

Prema Sweller, Aryes i Kalyuga (2011), razmatranje kognitivne arhitekture i organizacije pamćenja daje kontekst i objašnjenje zašto su neki postupci podučavanja efikasni, a neki ne.

U okviru TKO se razmatra koncept kognitivnog opterećenja kao opterećenje radne memorije prilikom procesiranja informacija. TKO razlikuje opterećenje nastalo (1) zbog prirode samog zadatka, koje naziva intrinzično opterećenje i (2) opterećenje koje je nastalo zbog same instrukcije koja se zadaje, koje naziva irelevantno opterećenje. U prvim verzijama teorije govorilo se samo o ove dvije vrste opterećenja. U drugoj polovini 1990-tih Sweller i sur. (1998) uvode, pored „neophodnog“ intrinzičnog opterećenja i „lošeg“ irelevantnog, i „dobro“, relevantno opterećenje, koje ovisi od intrinzičnog. Pretpostavka je da do relevantnog

opterećenja dolazi zbog stvaranja kognitivnih shema. Prema TKO, dizajn instrukcije treba da izaziva što manje irelevantno opterećenje i pri tome što veće relevantno opterećenje.

Pretpostavka u okviru TKO je da su intrinzično i irelevantno opterećenje aditivni. Ukoliko je intrinzično kognitivno opterećenje nisko, visoko irelevantno opterećenje ne mora biti problem, jer u ukupnom zbroju kognitivno opterećenje koje zadatak izaziva ne opterećuje radno pamćenje preko njegovog kapaciteta.

8.2.1. Intrinzično kognitivno opterećenje

Intrinzično kognitivno opterećenje je dodano u teorijski okvir nakon irelevantnog opterećenja. U početnim fazama razvoja TKO fokus je bio na redukciji irelevantnog opterećenja. Potreba za dodatnim objašnjenjem se javila nakon što su se u nekim varijacijama uobičajenih eksperimenata u ovoj oblasti pojavile izvjesne nedosljednosti. Postalo je jasno da objašnjavanje ovih nelogičnosti (anomalija) zahtijeva novi koncept, koji će uzeti u obzir kompleksnost zadatka ili nivo interaktivnosti elemenata (Kalyuga, 2011).

Intrinzično opterećenje proizilazi iz same strukture informacija koje se trebaju usvojiti. Interaktivnost elemenata dovodi do intrinzičnog kognitivnog opterećenja. Različiti sadržaji podrazumijevaju različit nivo interaktivnosti elemenata i samim tim imaju različito kognitivno opterećenje. Zadaci sa niskom interaktivnošću elemenata se sastoje od jedinične informacije koju je moguće razumjeti samostalno, dok se zadaci sa visokom interaktivnošću elemenata sastoje od više elemenata, koji se mogu razumjeti samo u odnosu na ostale elemente u zadatku. Primjer koji daju Sweller, Aryes i Kalyuga (2011) za sadržaj sa niskom interaktivnošću elemenata je učenje riječi stranog jezika, jer se pojedinačne riječi mogu razumjeti neovisno jedna od druge. S druge strane, jezička sintaksa je primjer zadatka sa visokom interaktivnošću, s obzirom na to da ju je moguće razumjeti jedino ukoliko se razumiju individualni elementi, kao i njihovi međuodnosi. Učenje simbola hemijskih elemenata je zadatak sa niskom interaktivnošću. Ovaj zadatak, kao i zadatak savladavanja rječnika drugog jezika je zahtjevan zbog količine informacija koja se treba usvojiti, ali ipak ne uzrokuje visoko kognitivno opterećenje. Kako u ovom zadatku nema elemenata u interakciji, nećemo govoriti o „razumijevanju“ ili „nerazumijevanju“, čak i ukoliko simbol nekog elementa zaboravimo. S druge strane, zadatak savladavanja sintakse jezika, ili hemijskih jednačina, posmatrajući samo broj elemenata s kojima se operiše, može se činiti i lakšim od prethodnih. Ipak, ovaj zadatak podrazumijeva visoku interaktivnost elemenata, kada rješavamo ovakve zadatke govorimo o „razumijevanju“ ili „nerazumijevanju“ u zavisnosti od toga da li smo uspjeli simultano obraditi sve elemente u interakciji. Zadaci sa visokom interaktivnošću elemenata izazivaju veće kognitivno opterećenje u odnosu na zadatke sa niskom interaktivnošću elemenata.

Sama analiza elemenata u interakciji nam ipak ne daje informacije o prirodi materijala kako ga doživljava svaki učenik. U zavisnosti od shema koje učenik već posjeduje, isti materijal može biti različito kompleksan. Ukoliko je set elemenata u interakciji uspješno uklopljen u postojeću shemu, onda je potrebno samo shemu procesirati u radnoj memoriji. U tom slučaju, kognitivno opterećenje je nisko. Na primjer, za osobu koja čita svoj maternji jezik svaka riječ se obrađuje kao shema (riječ se ovdje posmatra kao skup slova koja u

interakciji daju riječ) i čitanje postaje proces obrade više shema-riječi. Za nekoga ko tek uči taj jezik, svako slovo se posmatra kao element koji se treba obraditi u interakciji sa drugima, što predstavlja zadatak sa visokom interaktivnošću elemenata.

Intrinzično opterećenje ovisi od kompleksnosti zadataka, a zadaci su konstruirani kako bi ispunili određene obrazovne ciljeve. Iz ovoga proizilazi da intrinzično opterećenje zavisi i od obrazovnih ciljeva. Intrinzično opterećenje ovisi i od broja elemenata u interakciji, a broj elemenata u interakciji zavisi od prethodnog znanja. Dakle, intrinzično opterećenje zavisi i od prethodnog znanja učenika. Kada je postignuto poravnavanje težine zadatka sa prethodnim znanjem, instrukcija je primjerena i tada učeniku kognitivno opterećenje izaziva samo intrinzično opterećenje, koje njemu izaziva sam zadatak. U slučaju kada postoji nesuglasje između težine zadatka i nivoa predznanja, instrukcija izaziva irelevantno opterećenje za kognitivne resurse učenika.

8.2.2. Irelevantno kognitivno opterećenje

Irelevantno kognitivno opterećenje je centralni konstrukt za TKO. TKO je originalno nastala sa ciljem razvoja postupaka za smanjenje irelevantnog opterećenja prilikom poduke.

I ova kategorija opterećenja ovisi o interaktivnosti elemenata, ali za razliku od intrinzičnog opterećenja, ovdje je elemente u interakciji u potpunosti moguće modificirati instrukcijom (Plass, Moreno i Brünken, 2010). Instrukcija koja ne nameće irelevantno kognitivno opterećenje je u skladu sa principima efikasnog kognitivnog funkcioniranja.

Instrukcija koju karakterizira nepoticanje promjene u dugoročnoj memoriji, instrukcija koja pokušava da promijeni pohranjene informacije koristeći princip slučajnosti a ne princip prosuđivanja i reorganizacije, instrukcija koja generalno postoji na pretpostavci da se slučajno otkrivanje može podučavati, instrukcija koja ignoriše uske mogućnosti principa promjene time što ignoriše ograničenja radne memorije, najčešće je neučinkovita jer nameće nepotrebne elemente u interakciji (Plass, Moreno i Brünken, 2010). Na primjer, ukoliko učeniku zadamo zadatak da otkrivanjem savlada neki znanstveni princip, umjesto da se učeniku prezentira znanstveni zakon, daju mu se minimalne informacije na osnovu kojih je potrebno da se samostalno izvede pravilo. Ovaj postupak će vrlo vjerovatno postaviti velike zahtjeve na radno pamćenje, čime se narušava zakonitost principa promjene. Razmjena relevantnih informacija je minimalna, što narušava zakonitost principa posuđivanja i reorganizacije. Postupak savladavanja zakona počiva na slučajnosti (princip slučajnosti) i kompletan postupak (generiranje mogućeg rješenja i provjera tog rješenja) je vrlo spor, neučinkovit način prikupljanja informacija. Ovaj postupak podrazumijeva obradu jako puno elemenata koji nisu neophodni za učenje. Ovakav pristup za posljedicu ima minimalan napor za konkretno povećanje znanja u dugoročnoj memoriji, što je u potpunosti u kontradiktornosti sa generalnim ciljem svake poduke.

Dakle, prema TKO, irelevantno kognitivno opterećenje uzrokuje sam način prezentiranja sadržaja koji se podučava. Irelevantno opterećenje ne doprinosi učenju, tj. ne doprinosi stvaranju shema u dugoročnoj memoriji. Irelevantno opterećenje je pod utjecajem načina dizajniranja instrukcije. Dizajn instrukcije koji je u skladu sa principima kognitivnog

funkcioniranja dovodi do smanjenja irelevantnog opterećenja i time oslobađanja kognitivnog kapaciteta koji onda može biti iskorišten za učenje.

8.2.3. Relevantno kognitivno opterećenje

Relevantno kognitivno opterećenje također se može objasniti kroz interaktivnost elemenata, s tim da je njegova priroda drugačija od irelevantnog i intrinzičnog. Irelevantno i intrinzično opterećenje su određeni kombinacijom materijala i karakteristika učenika, s tim da je naglasak na samom materijalu, i karakteristike učenika su jedino bitne u kontekstu njegovog predznanja (Sweller, 2010). Za razliku od intrinzičnog i irelevantnog opterećenja, relevantno opterećenje je isključivo određeno karakteristikama učenika. Ono se odnosi na resurse radne memorije koje učenik ulaže u savladavanje intrinzičnog opterećenja. U tom smislu, relevantno opterećenje je neovisno od samog sadržaja ili načina prezentacije materijala.

Sweller (2010; 2011) i Kalyuga (2011) revidiraju početnu postavku relevantnog opterećenja kao treće aditivne vrste opterećenja. Prema trenutnim konceptualizacijama, relevantno opterećenje nema aditivni efekat sa ostale dvije vrste. Pod pretpostavkom da je nivo motivacije konstantan, učenik nema nikakvu kontrolu nad relevantnim opterećenjem. Ukoliko je intrinzično opterećenje visoko, relevantno opterećenje mora također biti visoko zbog toga što učenik mora uložiti veći dio svog kapaciteta radne memorije u savladavanje gradiva. U slučaju kada je irelevantno opterećenje povišeno, relevantno kognitivno opterećenje je umanjeno zbog toga što učenik troši kapacitet svoje radne memorije kako bi prevazišao teškoće koje nameće neadekvatna instrukcija. Dakle, relevantno opterećenje je ustvari samo funkcija kapaciteta radne memorije usmjerenih na elemente u interakciji (koji određuju intrinzično opterećenje). Što je više kapaciteta radne memorije utrošeno na irelevantno kognitivno opterećenje, to je manje dostupno za usmjerenje ka prevazilaženju intrinzičnog opterećenja. Ova formulacija podrazumijeva da je učenička motivacija visoka i da su svi dostupni kognitivni kapaciteti usmjereni na zadatak.

8.3. Postupci manipulacije kognitivnim opterećenjem

U okviru ranijih verzija TKO u fokusu je bilo irelevantno kognitivno opterećenje te „efekti kognitivnog opterećenja“, kako su nazvani postupci koji imaju za cilj smanjenje irelevantnog opterećenja. Kasnije, sa pomjeranjem fokusa istraživanja na intrinzično i relevantno, definirani su postupci koji imaju za cilj modifikaciju ove dvije vrste opterećenja u smjeru koji će poboljšati učenje. U nastavku je dat pregled postupaka za manipulaciju sve tri vrste kognitivnog opterećenja.

8.3.1. Postupci za manipulaciju intrinzičnim opterećenjem

Iako se u početku razvoja TKO pretpostavljalo da je intrinzičnim opterećenjem nemoguće manipulirati bez promjene samog zadatka, u novije vrijeme su otkriveni načini da se to postigne. Novija istraživanja iz područja TKO se upravo bave validacijom načina manipulacije intrinzičnim opterećenjem. Također, početna pretpostavka da je uvijek dobro smanjivati intrinzično kognitivno opterećenje nije se pokazala ispravnom. Schnotz i sur.

(2007) navode da snižavanje intrinzičnog kognitivnog opterećenja nije uvijek poželjno. Naime, intrinzično kognitivno opterećenje treba prilagođavati prethodnom znanju učenika, što implicira da je nekada potrebno povećati intrinzično opterećenje.

Schnotz i sur. (2007) navode kako je, da bi se stimuliralo efikasno učenje, instrukciju potrebno prilagoditi tako da su zadatak koji se uči i nivo predznanja učenika usklađeni. Intrinzično opterećenje ovisi od kompleksnosti zadataka, a zadaci su konstruirani kako bi ispunili određene obrazovne ciljeve. Iz ovoga proizilazi da intrinzično opterećenje zavisi i od obrazovnih ciljeva. Intrinzično opterećenje ovisi i od broja elemenata u interakciji, a broj elemenata u interakciji zavisi od prethodnog znanja. Dakle, intrinzično opterećenje zavisi i od prethodnog znanja učenika. Kada je postignuto usklađivanje težine zadatka sa prethodnim znanjem, instrukcija je primjerena i tada je učenikovo kognitivno opterećenje ustvari samo intrinzično opterećenje (koje izaziva sam zadatak). U slučaju kada postoji nesuglasje između težine zadatka i nivoa predznanja, instrukcija izaziva irelevantno opterećenje za kognitivne resurse učenika.

Također, Leppink i sur. (2015) navode da je za uspješno učenje neophodno da je intrinzično opterećenje optimalno (a ne minimalno, kako se prije govorilo). Naime, kako bi uopće došlo do učenja, zadatak mora kod učenika izazivati dovoljno kognitivnog opterećenja kako bi se aktivirao kognitivni potencijal i isti usmjerio na savladavanje zadatka. Drugim riječima, prelagani kao i preteški zadaci onemogućavaju učenje.

8.3.2. Postupci za manipulaciju irelevantnim opterećenjem

U okviru TKO opisano je šest postupaka manipulacije irelevantnim opterećenjem. Ovi postupci imaju za cilj smanjenje irelevantnog opterećenja s obzirom na to da visoko irelevantno opterećenje onemogućava učenje i prvobitno su opisani u početnoj fazi razvoja TKO (Sweler, Aryes i Kalyuga 2011). Prikaz ovih efekata dat je u Tabeli 8.1.

8.3.3. Postupci manipulacije relevantnim opterećenjem

Teorija kognitivnog opterećenja ima značajne implikacije za dizajn instrukcije prilikom učenja, koja bi, da bi bila efikasna, trebala držati irelevantno opterećenje na minimumu. Ipak, samo oslobađanje kognitivnog kapaciteta smanjenjem irelevantnog opterećenja nije dovoljno kako bi se instrukcija smatrala efikasnom. Pored smanjenja irelevantnog, potrebno je kapacitet radne memorije usmjeriti u savladavanje intrinzičnog opterećenja.

Neki od postupaka poticanja relevantnog opterećenja opisani u literaturi su: primjeri za vježbu, poticanje samoobjašnjenja, vođenje i pružanje povratne informacije.

a) Primjeri za vježbu

Paas i sur. (2003) navode da je tehnika davanja primjera za vježbu (inicijalno opisana kao tehnika smanjenja irelevantnog opterećenja) učenicima koristila kada je ustvari poticano relevantno opterećenje. Različiti primjeri potiču aktivno stvaranje shema jer potiču uočavanje i vremenom razlikovanje relevantnih od irelevantnih karakteristika zadatka.

Tabela 8.1. Efekti kognitivnog opterećenja fokusirani na smanjenje irelevantnog kognitivnog opterećenja

<i>Efekt kognitivnog opterećenja</i>	<i>Kratak opis</i>
Efekt slobodan od specifičnog cilja	Smanjuje irelevantno opterećenje time što fokusira pažnju učenika na karakteristike problema i sva dostupna rješenja. U toku poduke potrebno je zamijeniti konvencionalne probleme sa problemima u kojima nije definiran cilj.
Efekt urađenih i parcijalno urađenih primjera	Smanjuje irelevantno opterećenje fokusirajući pažnju učenika na strategije i korake u rješavanju problema. Postiže se zamjenom konvencionalnih problema sa potpuno ili djelomično urađenim primjerima koji se analiziraju.
Efekt podijeljene pažnje	Reorganizacija međusobno povezanih informacija iz različitih izvora na način da se nalaze na jednom mjestu, smanjuje irelevantno kognitivno opterećenje smanjujući napor neophodan za mentalnu integraciju informacija.
Efekt redundanse	Zamjenjivanje više različitih izvora informacija koje je moguće zasebno razumjeti sa jednim izvorom informacija smanjuje irelevantno kognitivno opterećenje time što eliminira obradu redundantnih informacija.
Efekt obrnute ekspertize	Smanjuje irelevantno kognitivno opterećenje prilagodbom stepena složenosti sadržaja učenikovom prethodnom znanju.
Efekt modaliteta	Smanjuje irelevantno opterećenje time što multimedijalna prezentacija sadržaja koristi vizualne i auditivne procesore radne memorije. Postiže se zamjenom napisanog teksta objašnjenja i nekog drugog vizualnog izvora informacija (npr. dijagram i objašnjenje) sa vizualnom prezentacijom i izgovorenim objašnjenjem.

Kako smo već naveli, fokus većine istraživanja TKO je pronalazak načina dobrog dizajna instrukcije kako bi se smanjilo irelevantno i maksimiziralo relevantno opterećenje. Često smo ipak suočeni sa situacijom u kojoj zadatak, čak i nakon eliminiranja svih izvora irelevantnog opterećenja, izaziva previše kognitivnog opterećenja. U ovim situacijama se nameće imperativ smanjenja intrinzičnog opterećenja kako bi se omogućilo učenje (tj. povećalo relevantno opterećenje).

Pass i sur. (2003) navode tehniku povećavanja relevantnog opterećenja pomoću povećanja varijabilnosti zadataka za vježbu. Ova tehnika podrazumjeva vježbanje više različitih verzija zadatka bez nekog posebnog reda. Moguće je mijenjati površinske karakteristike zadatka (kao što je kontekst) ili samu strukturu zadatka. Ovakav način učenja se nije pokazao uspješnim kada se mjeri učinak na lakšim zadacima, ali su ispitanici koji su ovako učili lakše zadatke postizali bolje rezultate na testovima retencije.

b) Poticanje samoobjašnjenja

Problem sa pristupom koji podrazumijeva radne primjere je u tome što često učenici ponuđene primjere ne procesiraju dubinski. Kao rezultat toga, učinkovitost ove metode opada. Jedan od načina povećanja dubine procesiranja je poticanje samoobjašnjenja. Većina učenika ne primjenjuje samoobjašnjenja, ili ukoliko ih primjenjuje ona su površna i ustvari podrazumijevaju ponavljanje prezentiranih informacija. Prema Clark i sur. (2006) istraživači su identificirali tri vrste strategija primjenjivih u produktivnim samoobjašnjenjima i to: 1. praćenje i ispravljanje, 2. pokušaj i provjera i 3. uočavanje veza rješenja sa općim principom koji je primijenjen.

c) Vođenje i pružanje povratne informacije

Tehnika koja se koristi za poticanje relevantnog opterećenja se odnosi na pružanje vođenja i davanja povratnih informacija prilikom podučavanja. Kod jednostavnih zadataka, smanjenje vođenja prilikom učenja obično ima pozitivan efekat na učenje. Naime, kod jednostavnih zadataka vođenje ima pozitivne efekte na učinak tokom samog učenja, ali kada se naknadno poredi učinak skupine koja je imala manje i skupine koja je imala više vođenja, druga skupina ima bolje rezultate. Slične nalaze dobijamo i kada izučavamo utjecaj povratne informacije na učinak. Istraživanja pokazuju da je bolje davati povratnu informaciju nakon većih koraka u učenju nego konstantno davati povratnu informaciju (prema Pass i sur., 2006).

S druge strane, istraživanja pokazuju potpuno drugačije rezultate ove tehnike prilikom podučavanja kompleksnih zadataka. Ovdje istraživanja pokazuju da pozitivne efekte vođenja i pružanja povratnih informacija prilikom podučavanja složenih zadataka.

Zaključak do kojeg dolaze Pass i sur. (2006) u svojoj analizi postupaka poticanja kognitivnog opterećenja je da je neophodno praviti razliku u podučavanju jednostavnih i kompleksnih zadataka, kako bi se poticali procesi relevantnog kognitivnog opterećenja. Oni navode da tehnike podučavanja kao što su vođenje i povratne informacije nisu učinkovite prilikom podučavanja kompleksnih zadataka te da ne potiču transfer u učenju, dok su se pokazale uspješnim u podučavanju lakših zadataka. Preporuka na osnovu urađenih istraživanja je da se prilikom podučavanja kompleksnih zadataka, kako bi se potaklo relevantno opterećenje, ustvari prvo pristupi nekoj od gore opisanih tehnika smanjenja intrinzičnog opterećenja kako bi se potaklo relevantno opterećenje. Istraživanja su pokazala da se ovakvim pristupom potiče relevantno opterećenje, o čemu zaključujemo na osnovu toga što ispitanici koji su ovako podučavani postižu bolje rezultate na testovima učinka i transfera u učenju.

Paas i sur. (2003) navode da je nekada irelevantno opterećenje neodvojivo povezano sa relevantnim. Ovdje je potrebno navesti da nije uvijek neophodno praviti intervencije koje imaju za cilj smanjivati opterećenje. Naime, ukoliko ukupno opterećenje ne prevazilazi kapacitete učenika, centralno pitanje nije količina opterećenja nego njegov izvor. Ukoliko su izvor kognitivnog opterećenja mentalne aktivnosti koje ometaju stvaranje shema i automatizaciju, onda je izvor opterećenja sama instrukcija i usvajanje novih sadržaja je otežano. S druge strane, ukoliko opterećenje proizilazi iz mentalnih aktivnosti, onda se radi o relevantnom opterećenju koje potiče učenje. Prilikom konstruiranja instrukcije potrebno je,

dakle, imati u vidu to da nije dovoljno smanjiti irelevantno opterećenje, nego je neophodno poticati relevantno opterećenje.

8.4. Teorija kognitivnog opterećenja i upotrebljivost: preklapanja

Analizirajući historijski razvoj TKO i koncepta upotrebljivosti možemo uočiti određene sličnosti. Kognitivne teorije kao što je recimo Geštalt psihologija, su utjecale i na TKO i ne upotrebljivost. Dalje, razvojem TKO fokus se pomjerio sa smanjenja irelevantnog kognitivnog opterećenja na razvijanje postupaka za optimizaciju intrinzičnog i povećanje relevantnog kognitivnog opterećenja. Ovi postupci nekad u totalu dovode do povećanja ukupnog opterećenja, ali istraživanja su pokazala da, ukoliko se realiziraju na ispravan način, pospješuju učenje.

U kontekstu upotrebljivosti, fokus se pomjerio samo sa razvijanja softvera koji će omogućiti završetak nekog zadatka na što brzi i efikasniji način, na razvoj softvera koji će biti poticajan za korištenje i kada nije nužno neophodno da se zadatak što prije izvrši.

Neki principi definirani u okviru TKO se već primjenjuju u dizajnu korisniku prilagođenog softvera. Ovo važi posebno za efekt podijeljene pažnje i efekt redundanse. Heuristika koja se koristi u dizajnu softvera govori o tome da je informacije koje su međusobno povezane i neophodne za završetak zadatka neophodno integrirati u jedan prozor, bez da korisnik mora da pamti informacije iz jednog dijela i da ih koristi u drugom dijelu. Također, u okviru upotrebljivosti se primjenjuje Geštalt princip prostorne i vremenske povezanosti informacija. I TKO i područje upotrebljivosti uzimaju u obzir individualne razlike tj. karakteristike korisnika/učenika, i to posebno predznaja prilikom dizajniranja instrukcije/softvera.

Hollender i sur. (2010) predlažu dva modela integracije konceptata TKO i upotrebljivosti. Ovi modeli su ustvari razmatranja konceptata na višem nivou, koje je potrebno konkretizirati u implikacije za istraživanja i dizajn softvera, posebno obrazovnog softvera. U prvom modelu, sam softver kao medij prenosa znanja se tretira kao vrsta irelevantnog kognitivnog opterećenja. Ovdje se akcentira važnost dizajniranja softvera koji se lako koristi i koji se lako uči kao i pružanja obuke o softveru što je prije moguće u okviru procesa učenja.

U drugom modelu su integrirani koncepti TKO u osnovne koncepte kojima se opisuje dizajn na korisnika usmjerenog softvera. Koncept relevantnog kognitivnog opterećenja je ključan u ovom modelu, u kojem se akcentira važnost dizajniranja adekvatnih softverskih alata za podršku uspješnom završavanju zadatka. Konačni cilj korištenja softverskog alata je izgradnja mentalnih shema, tj. integriranje novog sa postojećim znanjem kako bi se novo znanje trajno pohranilo i adekvatno koristilo. Dakle, dobar softver treba da potakne relevantno kognitivno opterećenje.

Anmarkrud, Andresen, Bråten, I. (2019) u komentaru i objašnjenju rezultata do kojih su došli navode kako značajan procenat istraživanja upotrebe tehnologije u nastavi, konkretno multimedije koristi upravo okvir koji je ponudila TKO. Logika koja se nalazi u pozadini se bazira na tome da se uspješnost učenja uz pomoć multimedije najbolje može objasniti kroz analizu prilagođenosti sadržaja i prezentacije informacijama sa načinom na koji funkcioniра

radno pamćenje i na koji se informacije obrađuju. Iz navedenog prema autorima proizilazi da u zavisnosti od toga kako je multimedija dizajnirana, može poticati učenje kroz poticanje relevantnog opterećenja, ili otežavati učenje povećanjem irelevantnog opterećenja zbog neadekvatnog dizajna.

Ovdje je potrebno posebnu pažnju posvetiti pitanju povratne informacije. Naime, povratna informacija, i to adekvatna, pravovremena i individualna se smatra jednim od najznačajnijih faktora za poticanje motivacije za učenje. Dalje, istraživanja su pokazala da upravo upotrebom povratne informacije potičemo relevantno i optimiziramo intrinzično opterećenje. Kako smo već objasnili, u okviru TKO je u u prvim fazama razvoja elaborirana i dokazana potreba za usklađivanjem instrukcije sa predznanjem učenika (efekat „obrnuti ekspert“), a u kasnijim se važnost upravo ovog poravnanja pokazala u svjetlu značaja optimizacije intrinzičnog opterećenja i povećanja relevantnog kognitivnog opterećenja. U nastavnim situacijama koje su dominantne u našim školama i centrima za obrazovanje, nastavnik podučava više od dvadeset učenika u jednoj vremenskoj tački, nemoguće je da prilikom predstavljanja sadržaja uzme u obzir sve individualne razlike u predznanju. Dalje, prilikom davanja povratne informacije, nerijetko se također dešava da učenik ne dobije detaljnu i adekvatnu informaciju o svojim greškama kao i uputu na koji način popraviti te greške. Ove navedene teškoće s kojima se susrećemo u učionici se upotrebom obrazovnog softvera mogu značajno smanjiti ako ne i eliminirati. Nove tehnologije kao što su konverzijski agenti i vještačka inteligencija omogućuju individualizaciju poduke, evaluaciju i općenito pružanje različitog spektra usluga na način na koji do sada nije bilo moguće. Istraživanja pokazuju da upotreba ovih tehnologija dovodi do boljih ishoda učenja, brže, jeftnije i efikasnije usluge (Anmarkrud, Andresen, Bråten, I, 2019). Principi identificirani u okviru TKO su u mnogome već inkorporirani u ove tehnologije i to zahvaljujući istraživanjima upotrebljivosti koja su došla do rezultata koji su kompatibilni sa TKO. Podučavanje u kojem se koriste ove tehnologije je moguće maksimalno individualizirati za svakog korisnika, ukoliko se u sam algoritam integriraju mjere predznanja i praćenje usvajanja novog znanja. Softver može vrlo lako ukazati na greške, učeniku ponoviti/dodatno proširiti dio gradiva s kojim ima teškoće i ponuditi dodatne strategije vježbe/učenja koje može koristiti. Ovakva individualizacija, tj. usklađivanje softvera sa individualnim razlikama korisnika je upravo jedan od značajnijih zadataka koji su definirani u okviru novih konceptualizacija upotrebljivosti i dizajna na korisnika usmjerenog softvera.

8.5. Teorija kognitivnog opterećenja i softveri u obrazovanju

Skulmowski i Man Xu (2021) daju pregled istraživanja vezanih za kognitivno opterećenje u digitalnom i online učenju. Autori ističu da novija istraživanja u vezi različitih faktora instrukcijskog dizajna, iz područja digitalnog i online učenja, dovode u pitanje pretpostavku irelevantnosti kognitivnog opterećenja vezanog za softvere u obrazovanju. Interaktivni mediji za učenje, imerzija, realizam i redundantni elementi predstavljaju te uočene izazove, budući da se pokazalo da ovi faktori dizajna izazivaju kognitivno opterećenje koje nije relevantno za zadatak, tj. dodatno opterećenje, dok podižu motivaciju i promiču učenje. Autori ističu da još ne postoji jedinstven pristup za integraciju identificiranih efekata u TKO. Uključujući aspekte konstruktivnog usklađivanja, pristupa koji ima za cilj podsticanje

dubokih oblika učenja u cilju postizanja specifičnih ishoda učenja, pred istraživačima je zadatak da osmisle strategiju za balansiranje kognitivnog opterećenja u digitalnom učenju.

Skulmowski i Man Xu (2021) smatraju da je najvažnije da se detaljno razmotre i pozitivni i negativni efekti na kognitivno opterećenje koje identificirani faktori dizajna digitalnog učenja mogu uzrokovati. Osim toga, brojni rezultati istraživanja ističu da se pozitivni efekti digitalnog učenja mogu otkriti samo pomoću odgovarajućih metoda vrednovanja znanja. Predložena strategija usklađivanja kognitivnog opterećenja sa željenim ishodima učenja bit će korisna kako za formuliranje istraživačkih hipoteza, ali autori smatraju da je posebno korisna za praktičare koji trebaju da prihvate nove tehnologije cjelovito u nastavnom procesu.

Naše iskustvo u evaluaciji upotrebljivosti programa za e-učenje mjerenjem kognitivnog opterećenja (Zukić i Bošković, 2019) vezano je za mjerenje intrinzičnog i irelevantnog kognitivnog opterećenja prilikom korištenja e-softvera za učenje. Na nastavi iz predmeta Informatika, ispitanici su dobili zadatak da riješe jednostavni matematički problem koristeći aplikaciju Geogebra. Zadatak je bio da se opiše kružnica oko trougla. S obzirom na matematičko predznanje koje učenici imaju u srednjoj školi, ovaj zadatak sa matematičkog aspekta nije trebao biti zahtjevan, ali smo u ispitivanje uključili i učenike prvog i četvrtog razreda, da bismo mogli uočiti eventualne razlike.

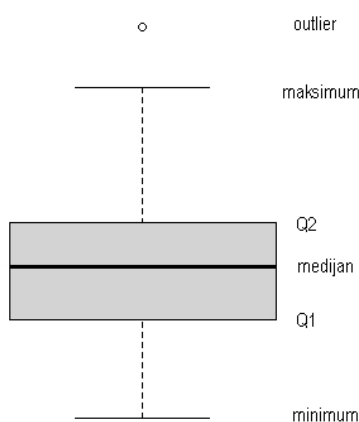
U ovoj studiji poređenje dvije grupe ispitanika je izvršeno t-testom i rezultati su ukazali da između dvije grupe učenika nema statistički značajne razlike na mjerama intrinzičnog i irelevantnog opterećenja. Vrijednost aritmetičke sredine za dvije vrste kognitivnog opterećenja u obje grupe su potvrdile naše inicijale pretpostavke, s obzirom da su obje grupe ispitanika doživjele nisko intrinzično opterećenje, tj. zadatak im je bio lagan, a s druge strane doživljeno je skoro jednako značajno irelevantno opterećenje, dakle ispitanici su snalaženje u softveru doživjeli teškim. U sklopu evaluacije kognitivnog opterećenja evaluirali smo i uspješnost rješavanja zadatka, pri čemu smo se suočili sa novim izazovom koji je direktno vezan za zaključak da se pozitivni efekti digitalnog učenja mogu otkriti samo pomoću odgovarajućih metoda vrednovanja znanja (Skulmowski i Man Xu, 2021). Tradicionalni pristup na papiru i korištenjem šestara za problem crtanja kružnice koja opisuje trougao uključuje: određivanje barem dvije simetrane stranice trougla, zatim određivanje mjesta presjeka tih simetrala što je centar kružnice, i crtanje same kružnice. Od 124 ispitanika njih 104 je riješilo zadatak, ali od toga su 32 ispitanika pronašla u Geogebri komandu za crtanje kružnice kroz tri tačke i tako dobili efikasnije, urednije i tehnologiji primjerenije rješenje. Poseban značaj navedenog istraživanja je zapravo u generisanju pitanja da li bi profesori ovo prihvatili kao riješen zadatak. Dodatno se potiču dalje diskusije vezanu za uticaj uvođenja tehnologije u nastavu, ne samo kao pomagala već da je potrebno sagledati obuhvat potrebnih promjena, kako načina podučavanja, ali i metoda i vrednovanja znanja.

9. Značaj eksplorativne analize prikupljenih podataka

John Tukey je prepoznao važnost pažljivog istraživanja podataka kako bi se otkrile strukture i informacije sadržane i skrivene u podacima, i uveo je i pojam *eksplorativne analize podataka* - EDA u svojoj istoimenoj knjizi *Explorative Data Analysis* (Tukey, 1977).

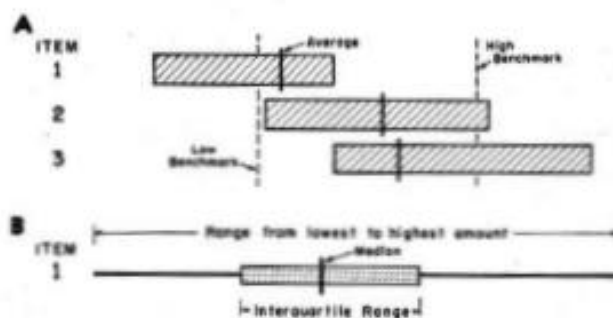
Eksplorativna (istraživačka) analiza podataka je pristup analizi skupova podataka s ciljem sumiranja njihovih glavnih karakteristika koristeći metode vizualizacije podataka. Ovo je važan prvi korak u analizi podataka kako kod statističkih analiza, tako i kod pripreme podataka za obradu metodama mašinskog učenja.

Tukey je promovirao korištenje petobrojnog sumiranja numeričkih podataka: dva ekstrema - maksimum i minimum, medijan i kvartili. Isticao je da su medijan i kvartil definirani za sve distribucije i da su robustniji za iskrivljene distribucije. U skladu s time, predstavio je i promovirao grafički metod korištenja pravougaonika u kombinaciji sa specifičnim tačkama koji prikazuje informacije o raspodjeli vrijednosti, kao kvartile, medijan, minimalnu i maksimalnu vrijednost - kutijasti plot (eng. *box plot*). Na slici 9.1. i je dat primjer kutijastog plotu sa karakterističnim elementima.



Slika 9.1. Šematske prikaz raspodjele vrijednosti (box-plot) i prikaz specifičnih tačaka na box-plotu i njihova značenja

Treba prepoznati da je prvi grafikon slične namjene još 1952. predstavila Mary Spear, nazivajući ga stupac opsega (eng. *range bar*). Originalni primjer iz njene knjige dat je na slici 9.1. (Spear, 1952). Na slici 9.2. pod A ukazuje se na olakšano poređenje grupa podataka. Na primjeru prikazanom na slici 9.2. A se porede tri grupe podataka, svaka predstavljena stupcem opsega i sa označenom srednjom vrijednosti, a na slici je vizualizirano i dodatno poređenje sa vrijednostima donji i gornji benchmark. Ista slika, pod B, daje uopšteni primjer grafa sa interkvartilnim vrijednostima prikazanim kao pravougaonik i sa označenom srednjom vrijednosti, uz dodatne linije koje ukazuju na čitav opseg.



Slika 9.2. Prvi grafikon šematske raspodjele vrijednosti - stupac opsega (Spear, 1952)

Grafički prikazi podataka su u osnovi eksplorativne analize podataka, ali se ne smiju iz fokusa izgubiti ciljevi ove analize:

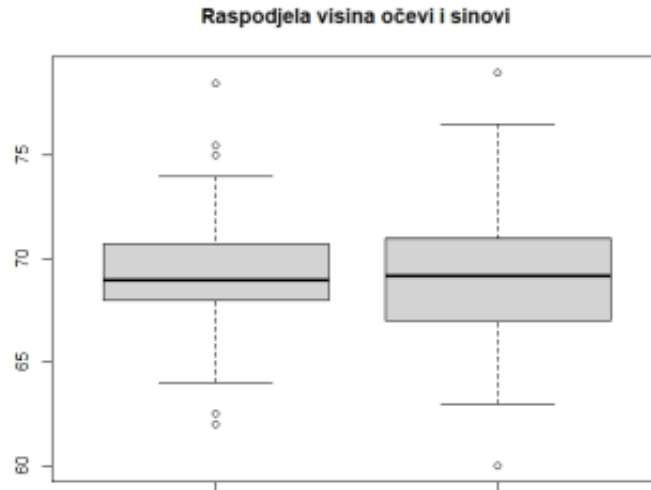
- omogućiti otkrića u podacima koja nisu očekivana, posebno moguće uzročne veze,
- predložiti hipoteze o uzrocima uočenih pojava,
- procijeniti zadovoljenje pretpostavki na kojima se trebaju zasnivati statistički testovi (npr. normalna raspodjela),
- uočiti greške u podacima, strčeće i nedostajuće vrijednosti,
- pomoći kod izbora odgovarajućih statističkih testova,
- sačiniti smjernice za dalje prikupljanje podataka putem anketa ili eksperimenata

Kod primjene metoda mašinskog učenja za analizu podataka radi se sa velikim količinama podataka, tako da je uočavanje grešaka u podacima i njihova korekcija značajan, mukotrpan i obiman zadatak, koji se naziva „čišćenje podataka“, i obično uzima više od 50% sveukupnog vremena. Kod primjera statističke obrade podataka koje prezentiramo u ovom poglavlju primarni fokus kod eksplorativne analize podataka je na metodama vizualizacije i njihovoj pomoći u otkrivanju uzoraka i veza među podacima.

Tukey je istaknuo da je najveća vrijednost grafikona kada nas natjera da primijetimo ono što nismo očekivali da ćemo vidjeti. Grafikoni imaju ulogu da nam ilustruju kvantitativne veličine, da nam ukažu na trendove promjene, da uočimo raspodjele i asocijacije koje postoje između veličina.

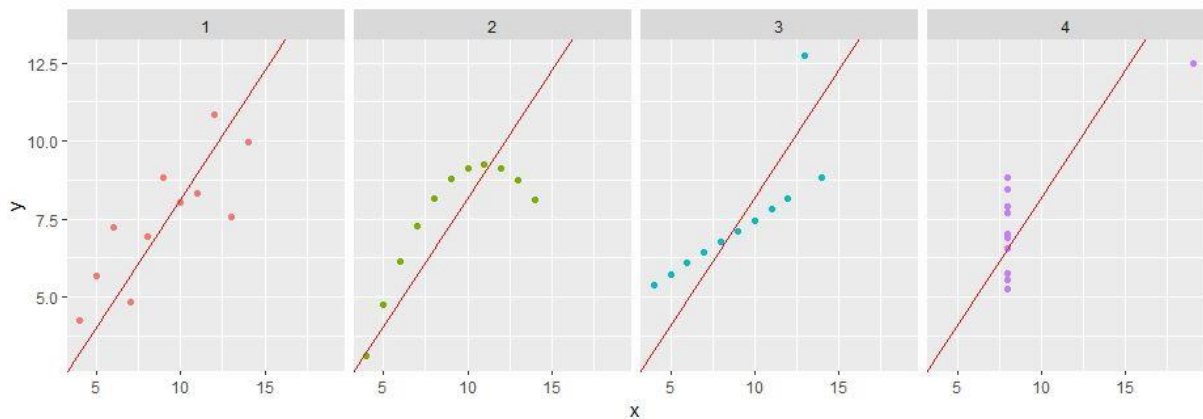
9.1. Primjeri vizualizacije kod eksplorativne analize podataka

Najčešći primjer korištenja kutijastih plotova je mogućnost da se na jednom grafikonu prikaže i poredi više grupa podataka. Primjer je dat na slici 9.3. gdje su prikazane raspodjele visina očeva i sinova iz Galtonovog seminalnog istraživanja (Galton, 1886).



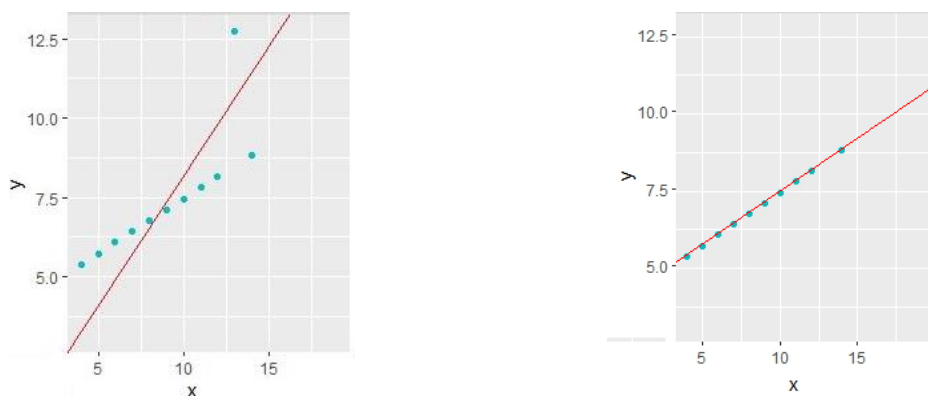
Slika 9.3. Poređenje grupa podataka pomoću dva box plota

Drugi primjer koji ćemo predstaviti je Anscombov kvartet. Francis Anscombe je svojim kvartetom grafikovana sredinom 1975. godine ukazao na poseban značaj vizualizacije podataka – svaki je graf prikazivao linearnu regresiju za skup od jedanaest tačaka. Svaki od četiri skupa imaju daju iste statističke vrijednosti za regresijsku pravu, dok vizuelno predstavljene izgledaju veoma različito i ukazuju na neophodne transformacije podataka prije statističkog modeliranja (Anscombe, 1973). Ova ideja ilustrirana na slici 9.4. je dovela u pitanje opšteprihvaćeni koncept da su samo numeričke kalkulacije precizne, dok su grafovi samo grube skice i nisu ozbiljan i naučan način prikazivanja statističkih podataka.



Slika 9.4. Grafovi Anscombeovog kvarteta

Kada želimo da utvrdimo da li postoji veza – korelacija - između opservacija predstavljenih dvijema promenljivim x i y , koristimo tačkasti raspšeni (*scatter*) dijagram. Na primjeru na slici 9.5. je dat dijagram jednog od skupova podataka iz Anscombeovog kvarteta (Anscombe 1973), i to skup koji ima linearnu vezu y i x uz jednu strčecu vrijednost (*outlier*) koja je uticala da linija regresije nema ispravan nagib. Ovo je ujedno primjer kako se vizualizacija koristi u eksplorativnoj fazi s ciljem eliminisanja strčecih vrijednosti iz skupa podataka i ponovnog modeliranja veze između x i y .



Slika 9.5. Tačkasti grafovi za veličine x i y koji ukazuju na njihovu korelaciju, prije i nakon uklanjanja strčćih vrijednosti (Boškovići sur., 2024)

9.2. Primjeri vizualizacije kod komunikacije rezultata

Vizualizacija podataka je značajna ne samo u eksplorativnoj fazi i već za komunikaciju dobijenih rezultata. Vizualizacije su značajne i kada rezultate prikazujemo u našoj grupi saradnika, ali i kod objavljivanja rezultata provedenih istraživanja. Prikazat ćemo ilustraciju iz primjera istraživanja Rizvić i sur. (2019) u kojem je korišten upitnik sa tri podskale mjerenja koje se odnose na: (A) značaj interaktivnog digitalnog pripovijedanja (eng. *Interactive Digital Storytelling*) - IDS za obrazovanje, (B) karakteristike IDS-a koje doprinose uspjehu u nastavi i (C) karakteristike IDS-a relevantne za motivaciju korisnika; sa ukupno 9 (A), 8 (B) i 4 (C) stavki Likertove skale sa 5 tačaka. Sadržaj upitnika je dat u tabeli 9.1.

Tabela 9.1. Primjer upitnik za istraživanje multimedijalnih softvera u nastavi (Rizvić i sur., 2019)

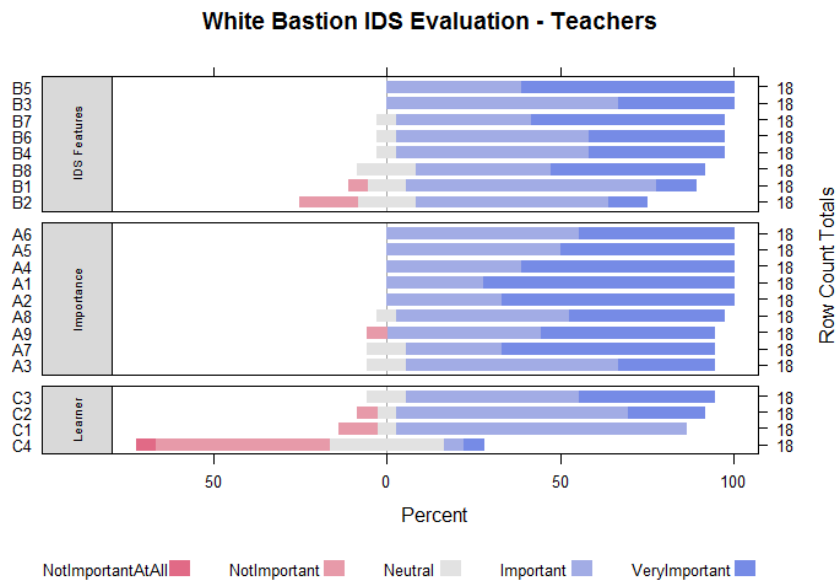
Oznaka stavke	Pitanje koje se ocjenjuje
A - Ocijenite koliko je prema vašem mišljenju važno sljedeće:	
A1	Korištenje multimedijalnih aplikacija u nastavi historije.
A2	Korištenje multimedijalnih aplikacija u podučavanju o umjetnosti.
A3	Učenje kroz interaktivne multimedijalne igre.
A4	Pristup kulturnom nasljeđu iz udaljenih destinacija.
A5	Pristup rekonstruisanom kulturnom nasljeđu.
A6	Motivisanje učenika da samostalno istražuju historiju.
A7	Motivisanje učenika za umjetnička djela.
A8	Uvezivanje znanja iz različitih disciplina.
A9	Promocija kulturnog nasljeđa u sklopu turističkog marketinga.

Tabela 9.1. Primjer upitnik za istraživanje multimedijalnih softvera u nastavi
(Rizvić i sur., 2019) (nastavak)

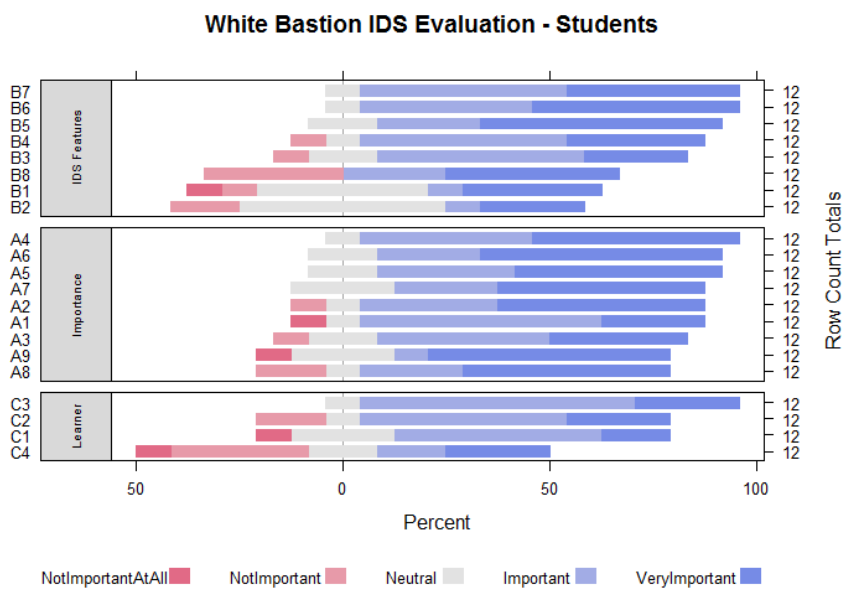
Oznaka stavke	Pitanje koje se ocjenjuje
B - Ocijenite koliko su po vašem osjećaju značajne sljedeće osobine za uspješnost multimedijalnog predstavljanja kulturnog nasljeđa:	
B1	Mogućnost upravljanja tokom izvršavanja aplikacije
B2	Postojanje zadatka kojeg korisnik ispunjava
B3	Pomoć u snalaženju u aplikaciji
B4	Osjećaj stvarnog prisustva uz objekat.
B5	Kvalitet tehničke izvedbe multimedijalne aplikacije
B6	Estetska izvedba multimedijalne aplikacije
B7	Vjernost original
B8	Inovativnost u multimedijalnom predstavljanju
C - Ocijenite uticaj sljedećih faktora na stepen uključenosti korisnika aplikacija:	
C1	Umijeće korisnika u radu sa multimedijalnim aplikacijama
C2	Vrijeme potrebno da se aplikacija pregleda
C3	Važnost aplikacije za korisnikovo učenje/rad
C4	Prijašnje znanje o objektu kulturnog nasljeđa

Na slici 9.6. su dati grafikoni koji ilustruju odgovore i to pod (a) odgovore profesora, a pod (b) učenika i studenata. Korišteni su divergentni naslagani stupičasti grafikoni, što znači da su frekvencije svih odgovora naslagane na jedan stubac (vidite razliku prikazivanja frekvencija odgovora na slici 6.1.), a da su horizontalni stupci centrirani na dijelu neutralnih odgovora – zato se nazivaju divergentni. Odgovori su grupisani po podskalama A, B i C, a za izdvojene grupe korisnika kreirani su odvojeni grafikoni zbog poređenja.

Ovaj način prikazivanja omogućava jednostavan i direktan uvid u raspodjelu odgovora. Odmah se može uočiti da su učenici/studenti za podskale A i B kod više stavki bili neutralni ili su ocijenili da nemaju značaj, što nije slučaj sa profesorima. Veći je stepen slaganja u odgovorima za stavke podskale C kod profesora i studenata. Grafikoni dati u primjeru su isprogramirani u jeziku R.



(a)



(b)

Slika 9.6. Divergentni stupičasti grafikoni za ilustraciju odgovora na Likertovoj skali od 5 tačaka: (a) profesori i (b) studenti (Rizvić i sur., 2019)

Kako su alati za vizualizaciju danas dostupni, npr. MS Excel, Google Sheets, važno je savladati osnovne tehnike i principe vizualizacije i koristiti ih kod prezentacije rezultata.

Zaključak

Ideja vodilja autora prilikom pisanja ove knjige je da naši studenti nastave čitati i koristiti knjigu i kada postanu profesionalci u svom poslu. Cilj nam je da knjiga ima praktičnu vrijednost i da bude jedna od onih koja je „uvijek pri ruci“ kada se treba riješiti problem. Problem o kojem govorimo, koji nas muči i koji muči naše kolege, kako stručnjake u području razvoja softvera tako i psihologe koji se bave kognitivnom psihologijom, je kako napraviti softver, proizvod koji će korisnici sa lakoćom koristiti. Softver koji se nudi treba biti što je moguće više usklađen sa korisničkim prirodnim načinom rješavanja problema, pokretima i percepcijom. Ovo je posebno važno kada se softver koristi u obrazovnom procesu i kada su očekivanja da proces učenja olakša a ne usloži. Kako bi se to postiglo potrebno je uvidjeti koji je to prirodni način na koji se mi ponašamo u kontekstu u kojem se nudi softversko rješenje, te tehnički onda izraditi i ponuditi proizvod koji je s tim usklađen.

Za adekvatno identificiranje osnovnih karakteristika korisničkog interfejsa koji se dizajnira, potrebno je prikupiti podatke od osoba za koje očekujemo da će biti korisnici. Ovom koraku je neophodno pristupiti ozbiljno, znanstveno i studiozno jer je ključan za planiranje svih resursa potrebnih za proizvodnju. Prvi korak u ovom procesu je identificiranje korisničke ciljne grupe. Nakon toga prikupljamo adekvatan uzorak pojedinaca iz populacije od interesa. U trećem koraku se odlučujemo za istraživački pristup i instrumentarij koji ćemo primijeniti. Prikupljene podatke je potrebno adekvatno statistički obraditi i donijeti statističke zaključke. Na kraju, na osnovu statističkih zaključaka, uvida u kompletan proces, donosimo i praktične zaključke o našem proizvodu. Nekada se ovi koraci mogu primjenjivati prije dizajna proizvoda, a nekada istraživanje sprovodimo u toku razvoja proizvoda, na nekoj od verzija koju razvijamo.

U ovom udžbeniku su sažete osnovne informacije koje stručnjaci koji rade na razvoju softverskih rješenja treba da znaju prilikom planiranja istraživanja upotrebljivosti tj svakog od gore navedenih koraka. Cilj nam je bio ponuditi konkretne primjere, konkretna rješenja i razmatati relevantne praktične implikacije, bez prema našem sudu za ovu svrhu, nepotrebnih teorijskih razmatranja.

Industrija IT sektora u Bosni i Hercegovini je već neko vrijeme jedan od najbrže rastućih segmenata privrede. Vrlo često se od menadžera u IT sektoru čuje potreba za većom suradnjom sa obrazovnim institucijama, najviše univerzitetom, kako bi se odgovorilo na rastuće potrebe tržišta rada. Naša je namjera, da našim radom i doprinosom općenito, pa i ovim udžbenikom, ojačamo kompetencije naših trenutnih studenata, a budućih stručnjaka, za izlazak na tržište rada. Cilj nam je da ponudimo ne samo konkretna znanja i alate, nego i da razvijemo interdisciplinarni način razmišljanja i divergentno razmišljanje u rješavanju problema. Spoj psihologije i IT znanosti se možda na prvu ne čini kao logičan za postizanje ovih ciljeva, ali nakon dodatnog razmišljanja i čitanja, vjerujemo da će se korisnik ovog udžbenika, kao i mi same, zaljubiti u ovo područje i sve ono što su praktične implikacije istog.

Literatura

- Alshammari, S. H., & Rosli, M. S. (2020). A review of technology acceptance models and theories. *Innovative Teaching and Learning Journal (ITLJ)*, 4(2), 12-22.
- Anmarkrud, Ø., Andresen, A., & Bråten, I. (2019). Cognitive load and working memory in multimedia learning: Conceptual and measurement issues. *Educational Psychologist*, 54(2), 61-83.
- Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician*, Vol. 27 (1), pp. 17–21.
- Barnum, C. M. (2010). *Usability testing essentials*. Burlington, MA: Morgan Kaufmann.
- Bloom, B. S. (1974). Time and learning. *American Psychologist*, 29(9), 682–688. <http://dx.doi.org/10.1037/h0037632>
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, 6(1), 3–5. <http://www.jstor.org/stable/41613414>
- Clark C.R, Nguyenn F, Swellern (2007): *Efficiency in Learning: Evidence-Based Guidelines to Manage Cognitive Load*, San Francisko: John Wiley & Sons
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Curr Dir Psychol Sci*, 19, 51-57.
- Dai, D. Y., & Sternberg, R. J. (Ur.). (2004). *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development*. Routledge.
- Đapo, N. i Đokić, R. (2012). *Statistika u psihologiji Priručnik za studente*.
- Davis, F.D. i Granić, A. (2024). *The Technology Acceptance Model: 30 Years of TAM*. (n.p.): Springer International Publishing.
- Davis, F. D. (1986). A technology acceptance model for empirically testing new end-user information systems: Theory and results. Doctoral dissertation. MIT Sloan School of Management. <https://dspace.mit.edu/handle/1721.1/15192>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://www.jstor.org/stable/249008>
- De Winter, J. C. I Dodou, D. (2010). Five-point Likert items: t test versus Mann-Whitney-Wilcoxon. *Practical assessment, research & evaluation*, 15(11), 1-12.
- Duran, R., Zavgorodniaia, A., Sorva, J. (2022). Cognitive load theory in computing education research: A review. *ACM Transactions on Computing Education (TOCE)*, 22(4), 1-27.
- Fajgelj, S. (2005). *Metode istraživanja ponašanja*. Centar za primenjenu psihologiju.

- Galton, F. (1886), "Regression Towards Mediocrity in Hereditary Stature," *Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 246-263.
- Heiberger, R.M., i N.B. Robbins, "Design of diverging stacked bar charts for Likert scales and other applications," *Journal of Statistical Software*, vol. 57, no. 5, 2014, pp. 1-32.
- Hollender, N., Hofmann, C., Deneke, M., Schmitz, B. (2010). Integrating cognitive load theory and concepts of human–computer interaction. *Computers in human behavior*, 26(6), 1278-1288.
- Husremović, D. (2016). Osnove psihometrije za studente psihologije. *Filozofski fakultet Univerziteta u Sarajevu, Sarajevo*
- Kalyuga, S. (2011). *Cognitive load factors in instructional design for advanced learners*. Nova Science Publishers.
- Kovanovic, V., Gašević, D., Dawson, S., Joksimovic, S., Baker, R. (2016). Does Time-on-task Estimation Matter? Implications on Validity of Learning Analytics Findings. *Journal of Learning Analytics*, 2(3), 81-110. <https://doi.org/10.18608/jla.2015.23.6>
- Kuhail, M. A., Alturki, N., Alramlawi, S., Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973-1018.
- Lai, P. C. (2017). The literature review of technology adoption models and theories for the novelty technology. *JISTEM-Journal of Information Systems and Technology Management*, 14, 21-38.
- Lazar J., Feng J., i Hochheiser H. (2010). *Research Methods in Human-Computer Interaction*, prvo izdanje, John Wiley & Sons
- Lazar J., Feng J., i Hochheiser H. (2017). *Research Methods in Human-Computer Interaction*, drugo izdanje, Morgan Kaufmann
- Leppink, J. i Van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspect Med Educ.*, 4, 119-127.
- Martins, A.I., Rosa, A.F., Queirós, A. Silva, A. Pacheco Rocha, N. (2015). European Portuguese Validation of the System Usability Scale (SUS), *Procedia Computer Science*, Volume 67, Pages 293-300, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2015.09.273>.
- Marzuki, M., Fadhil, M. I sur. (2018). "Translation, Cross-Cultural Adaptation, and Validation of the Malay Version of the System Usability Scale Questionnaire for the Assessment of Mobile Apps." *JMIR human factors* vol. 5,2 e10308. 14 May. 2018, <https://doi:10.2196/10308>
- Milas, G. (2005). *Istraživačke metode u psihologiji i drugim društvenim znanostima*. Naklada slap.
- Miller, G. A. (1956). The magical number seven, plus or minus two. Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Nielsen J. (2012) . Usability 101: Introduction to Usability, Nielsen Norman Group, online <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>

- Paas, F., Renkl, A., Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational psychologist*, 38(1), 1-4.
- Perneger, T., Smith, G. S. (1991). The driver's role in fatal two-car crashes: a paired "case-control" study. *American journal of epidemiology*, 134(10), 1138–1145. <https://doi.org/10.1093/oxfordjournals.aje.a116017>
- Plass, J. L., Moreno, R., Brünken, R. (Ur.). (2010). *Cognitive load theory*. Cambridge University Press. Research Methods, 45, pristupljeno: 23.09.2023.
- Rizvic S, Boskovic D, Okanovic V, Sljivo S, Zukic M. (2019). Interactive digital storytelling: bringing cultural heritage in a classroom. *Journal of Computers in Education*. Mar 4;6:143-66.
- Rogovin, M. i Frampton, Jr., G. T., 1979. Three Mile Island: a Report to the Commissioners and to the Public. Volume I.
- Sauro, J. (2011). *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices*. Measuring Usability LLC
- Sauro, J., i Lewis, J. R. (2016). *Quantifying the user experience: Practical statistics for user research*. Drugo izdanje, Morgan Kaufmann.
- Sauro, J. (2018). 15 Common Rating Scales Explained, MeasuringU, online, dostupno na: <https://measuringu.com/rating-scales/>
- Shneiderman, B. (2011). Claiming success, charting the future: micro-HCI and macro-HCI. *interactions*, 18(5), 10-11.
- Skulmowski, A., i Xu, K. M. (2022). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational psychology review*, 34(1), 171-196.
- Spear, M.E. (1952). *Charting statistics*. McGraw-Hill
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677–680. <https://doi.org/10.1126/science.103.2684.677>
- Sweller, J., Ayres, P., Kalyga, S. (2011). *Cognitive load theory*. New York: Springer.
- Tlili, A., Shehata, B., Adarkwah, M. A. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *SmartLearn. environ.* 10, 15.
- Tukey, J.W. (1977). *Exploratory Data Analysis*. Addison-Wesley
- Van Merriënboer, J. J., i Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical education*, 44(1), 85-93.
- Van Nimwegen, C., Van Oostendorp, H., Burgos, D., Koper, R. (2006). Does an interface with less assistance provoke more thoughtful behavior? In 7th international conference on learning sciences (pp. 785–791). Bloomington, Indiana: International Society of the Learning Sciences

Venkatesh, V., i Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2), 186-204.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

Wu, C. H., Liu, C. H., Huang, Y. M. (2022). The exploration of continuous learning intention in STEAM education through attitude, motivation, and cognitive load. *International Journal of STEM Education*, 9(1), 1-22.

Zukić, M., i Bošković, D. (2020). Evaluacija upotrebljivosti programa za e-učenje mjerenjem kognitivnog opterećenja. Department of Pedagogy Book of Proceedings/Zbornik Radova Odsjeka za Pedagogiju, 3(3). Univerzitet u Sarajevu – Filozofski fakultet

