SVEUČILIŠTE U ZAGREBU
HRVATSKI STUDIJI

Ivan Flis

**INTERPRETIRAJU LI STUDENTI
PSIHOLOGIJE TESTOVE ZNAČAJNOSTI
NUL HIPOTEZA DOVOLJNO KRITIČNO?**
DIPLOMSKI RAD

Zagreb, 2013.

SVEUČILIŠTE U ZAGREBU
HRVATSKI STUDIJI

Ivan Flis

**INTERPRETIRAJU LI STUDENTI
PSIHOLOGIJE TESTOVE ZNAČAJNOSTI
NUL HIPOTEZA DOVOLJNO KRITIČNO?**

DIPLOMSKI RAD

Mentor:

Doc. dr. Iva Šverko

Zagreb, 2013.

SAŽETAK NA HRVATSKOM JEZIKU

INTERPRETIRAJU LI STUDENTI PSIHOLOGIJE TESTOVE ZNAČAJNOSTI NUL HIPOTEZA DOVOLJNO KRITIČNO?

Cilj istraživanja bilo je deskriptivno opisati koliko studenti psihologije u Hrvatskoj razumiju smisao statističkog testiranja značajnosti nul hipoteza s obzirom na uvriježeno shvaćanje u udžbenicima, koje je podložno bayesijanskoj i interpretativnoj kritici. Osim toga, rad predstavlja kratki pregled rasprava o smislu testova značajnosti i praksama poučavanja studenata. U istraživanju je sudjelovalo 350 studenata preddiplomskih i diplomskih studija psihologije sa pet fakulteta u Hrvatskoj (Zagreb – Hrvatski studiji i Filozofski fakultet, Rijeka, Zadar, Osijek). Uz to, cilj je bilo ispitati može li se razumijevanje testiranja nul hipoteza predvidjeti ocjenama, stavovima i interesima studenata. Razumijevanje testiranja nul hipoteza ispitano je Testom pogrešnog shvaćanja testova statističke značajnosti (NHST test) (Oakes, 1986; Haller i Krauss, 2002). Stavovi prema testiranju nul hipoteza su ispitani upitnikom stavova prema testiranju nul hipoteza konstruiranim za ovo istraživanje. Ocjene su operacionalizirane kao prosjek ocjena na preddiplomskog studiju iz svih kolegija, te uz njega prosjek ocjena na preddiplomskom i diplomskom studiju iz statističkih i metodoloških kolegija koje su studenti odslušali. Studenti su pokazali ograničeno razumijevanje na testu – postotak točnih odgovora na NHST testu nije prešao 56% ni na jednoj od šest čestica. Uz to, sustavno su pokazali lošije rezultate od njemačkih studenata u istraživanju Hallera i Kraussa (2002). Niti jedna varijabla – opći prosjek ocjena, prosjek ocjena iz metodoloških kolegija, pad barem jednog metodološkog kolegija, dvije varijable stava prema testiranju nul hipoteza te varijabla glavnog područja interesa u psihologiji – se nije pokazala prediktivnom u predviđanju točnosti odgovaranja na česticama NHST testa. Zaključak rada je da je potrebno razmotriti prakse poučavanja smisla i interpretacije testova značajnosti na hrvatskim katedrama psihologije.

Ključne riječi

Testovi značajnosti nul hipoteza, interpretacija p -vrijednosti, statističko obrazovanje psihologa u Hrvatskoj, kritike nul rituala.

**DO PSYCHOLOGY STUDENTS INTERPRET NULL HYPOTHESIS SIGNIFICANCE
TESTING CRITICALLY?**

The goal of the study was to descriptively analyze the understanding of null hypothesis significance testing among Croatian psychology students considering how it is usually understood in textbooks, which is subject to Bayesian and interpretative criticism. Also, the thesis represents a short overview of the discussions on the meaning of significance testing and how it is taught to students. There were 350 participants from undergraduate and graduate programs at five faculties in Croatia (Zagreb – Centre for Croatian Studies and Faculty of Humanities and Social Sciences, Rijeka, Zadar, Osijek). Another goal was to ascertain if the understanding of null hypothesis testing among psychology students can be predicted by their grades, attitudes and interests. The level of understanding of null hypothesis testing was measured by the Test of statistical significance misinterpretations (NHST test) (Oakes, 1986; Haller and Krauss, 2002). The attitudes toward null hypothesis significance testing were measured by a questionnaire that was constructed for this study. The grades were operationalized as the grade average of courses taken during undergraduate studies, and as a separate grade average of methodological courses taken during undergraduate and graduate studies. The students have shown limited understanding of null hypothesis testing – the percentage of correct answers in the NHST test was not higher than 56% for any of the six items. Croatian students have also shown less understanding on each item when compared to the German students in Haller and Krauss's (2002) study. None of the variables – general grade average, average in the methodological courses, two variables measuring the attitude toward null hypothesis significance testing, failing at least one methodological course, and the variable of main interest in psychology – were predictive for the odds of answering the items in the NHST test correctly. The conclusion of the study is that education practices in teaching students the meaning and interpretation of null hypothesis significance testing have to be taken under consideration at Croatian psychology departments.

Keywords

Null hypothesis significance testing, p-value interpretation, statistical education of psychologists in Croatia, null ritual criticism.

Sadržaj

1. UVOD	2
1.1 Post-pozitivistička shvaćanja psihologijske metodologije.....	2
1.2 Pogrešno shvaćanje testova značajnosti u psihologiji	5
<i>1.2.1 Povijesni pregled razvoja NHST</i>	<i>5</i>
<i>1.2.2 Kritika statističke osnove testova značajnosti u psihologiji</i>	<i>8</i>
<i>1.2.3 Kritika interpretacije testova značajnosti u psihologiji</i>	<i>10</i>
1.3 U obranu NHST.....	13
1.4 Kako poučavati testiranje značajnosti hipoteza?	16
1.5 Poučavanje testiranja značajnosti nul hipoteza – što je sa studentima?	17
3. METODA	22
3.1 Sudionici.....	22
3.2 Instrumenti.....	23
<i>3.2.1 Test pogrešnog shvaćanja testova statističke značajnosti (NHST test)</i>	<i>23</i>
<i>3.2.2 Stavovi prema testiranju nul hipoteza.</i>	<i>26</i>
<i>3.2.3 Upitnik demografskih i akademskih podataka.....</i>	<i>28</i>
3.3 Postupak.....	29
4. REZULTATI	30
4.1 Uspješnost u studiju: prosjeci ocjena općih i metodoloških kolegija.....	30
4.2 Interesi u psihologiji.....	31
4.3 Stavovi prema statističkom testiranju nul hipoteza	32
<i>4.3.1 Zadatak rangiranja.....</i>	<i>32</i>
<i>4.3.2 Upitnik stavova.....</i>	<i>33</i>
4.4 Razumijevanje testiranja nul hipoteza	34
4.5 Povezanost razumijevanja testiranja nul hipoteza s interesima, stavovima i ocjenama.....	35
5. RASPRAVA	40
6. ZAKLJUČAK.....	45
7. LITERATURA	46

1. UVOD

1.1 Post-pozitivistička shvaćanja psihologijske metodologije

Post-pozitivistička filozofija znanosti se već desetljećima udaljava od koncepcije objektivne znanosti, odnosno objektivne znanstvene metodologije u pozitivističkom¹ smislu. Feyerabend (1993), među ostalima, kaže: ne postoji teorijski neutralna znanstvena metoda. Svaka znanstvena metodologija je opterećena teorijom, i kako bi se približila nekom idealu objektivnosti, ta teorijska opterećenost se mora uzeti u obzir. Što se zaista time misli, posebno u slučaju psihologije?

Dobar primjer ovakvog razumijevanja znanstvene metodologije su povijesna istraživanja Kurta Danziger (1990) o tome kako se istraživački subjekt psihologije mijenjao kroz povijest discipline. Danzigerova post-pozitivistička (i konstruktivistička) teza je slijedeća: metodologija koju koristimo kako bi istraživali neku pojavu konstruira sam subjekt istraživanja. Metoda istraživanja ne daje samo odgovore na teorijski motivirana pitanja, već i konstruira mogući univerzum prihvatljivih pitanja koja se mogu postaviti. Danziger tvrdi da „takve sheme [metodološki postupci koje znanstvenici koriste] nisu samo kognitivni okviri za *interpretaciju* empirijskih podataka nego uključuju i praktična pravila za *produkciju* takvih podataka“ (Danziger, 1990, str. 3).

Dakle, ne postoji neki čisti pozitivistički krug u istraživanjima, gdje se teorija konstruira pa zatim testira empirijskim podacima, odnosno induktivno-deduktivna spirala kako je Milas naziva (2005, str. 44), jer su empirijski podaci koje koristimo kako bi odgovorili na teoretski motivirana pitanja (hipoteze) već teorijski obojeni. Kako bismo generirali empirijske podatke za svoja istraživanja, već prihvaćamo određene teoretske odrednice pri samom postupku produkcije podataka. U psihologiji je to već očito na terminološkoj razini, pošto znanstvenici koriste nešto što nazivaju teorijama mjerenja pri produkciji podataka koje koriste u svojem radu.

Na implicitnoj razini, u psihologiji je jasno da su podaci isto produkt teorije kao i empirijskog istraživanja, no to se rijetko eksplicitno naglašava. Vjerojatno i zbog brige o

¹ Logički pozitivizam je filozofski projekt shvaćanja znanosti razvijen početkom 20. stoljeća u takozvanom Bečkom krugu. Ovo shvaćanje znanosti se u suštini temeljilo na hipotetičko-deduktivnom modelu: znanstveni zakoni (univerzalni ili statistički) su hipoteze koje se testiraju njihovim deduktivnim posljedicama (empirijskim podacima). Ovakvo shvaćanje znanosti se uglavnom danas smatra idealističnim i prevaziđeno je kasnijim razvojem filozofije znanosti (Machamer, 2002).

statusu same discipline te čuvanju vlastitog identiteta i integriteta kao znanstvenika (Flis, 2012). Dovođenje u pitanje objektivnih temelja metode se često izjednačava sa dovođenjem u pitanje temelja discipline kao takve.

Pri razmatranju teorijske opterećenosti metodologije moramo se ograničiti na, u današnjoj psihologiji, najčešće metode istraživanja. Danas je to uvjerljivo kvantitativna metodologija - neki istraživači idu toliko daleko te kažu kako postoji kvantitativni imperativ u psihologiji (Michell, 2003). Kvantitativna metodologija, iz perspektive moderne psihologije, ima najveću znanstvenu legitimnost: podaci koji se generiraju pomoću nje imaju najveću percipiranu empirijsku vrijednost. U sociološkom smislu, kvantitativna metodologija psihologiji daje znanstveni legitimitet.

Danziger (1990, str. 68) ovaj zaokret u povijesti psihologije naziva „trijumfom agregata“. Koristeći riječ zaokret implicira se da dominacija kvantitativne metodologije nije povijesna činjenica od samog početka razvoja discipline, već kasniji razvoj. Danziger njegov povijesni početak smješta s Galtonovim i Pearsonovim laboratorijem za antropometriju te razvojem pripadajućih statističkih metoda za obradu podataka koji su se pojavili u njihovim istraživanjima.

Izvrstan primjer koji približava povijesnu činjenicu uspostave dominacije kvantitativne metodologije u psihologiji je rad u kojem Spearman (1904, str. 72-73) izražava oduševljenje sa (tada) novim konceptom korelacije, koji je kasnije postao osnova mnogih kompleksnijih statističkih postupaka (pošto je u osnovi generalnih linearnih modela, faktorske analize, strukturalnog modeliranja, itd.). Spearman naziva sve prijašnje pokušaje „procjene dokaza i preciznog mjerenje tendencija“ u psihologiji „naivnim neznanjem“ te nadalje tvrdi kako svi „naporni eksperimenti“ koje su psiholozi proveli i objavili do tada nisu nimalo napredniji u postupcima mjerenja nego ono što bi laik mogao napraviti. Prema njegovu viđenju, ono što razlikuje psihologa od laika je baš kvantitativna metodologija i pripadajući statistički postupci.

Stotinu godina nakon Spearmanovog oduševljenja, psihologijski udžbenici koji poučavaju studente metodama te mlade znanosti dekontekstualiziraju ovaj povijesni razvoj metodologije, i nude idealiziranu pozitivističku sliku „objektivne“ metodologije koja osigurava dolazak do „objektivnog“ znanja. Kroz urušavanje povijesnog konteksta, znanstvenoj disciplini se daje njen legitimitet. Shvaćena na taj način, ova dekontekstualizacija

je prirodna posljedica etabliranja psihologije kao znanosti. Kuhn (2002) takve udžbenike opisuje kao „sredstva za indoktrinaciju budućih stručnjaka u disciplinu“ ovim riječima:

Udžbenici tako počinju s dokidanjem znanstvenikova smisla za povijest vlastite discipline, a potom nastavljaju pružajući joj zamjenu za ono što su eliminirali. Karakteristično je da udžbenici znanosti sadrže samo malo povijesti, ili u uvodnom poglavlju, ili, još češće, u razasutim pozivanjima na velike heroje ranijih razdoblja. Temeljem tih pozivanja, i studenti i stručnjaci počinju se osjećati sudionicima jedne dugotrajne povijesne tradicije. No, ta tradicija koja je izvedena iz udžbenika i u kojoj znanstvenici počinju osjećati svoje sudjelovanje, tradicija je koja zapravo nikada nije niti postojala. Iz razloga koji su istodobno očiti i visoko funkcionalni, znanstveni udžbenici, kao i najveći dio starijih povijesti znanosti, pozivaju se samo na onaj dio rada ranijih znanstvenika koji se lako može promatrati kao doprinos postavljanju i rješavanju paradigmatičkih problema kojima se bavi. Djelomično putem selekcije, a djelomično putem iskrivljavanja, znanstvenici ranijih vremena implicitno su predstavljeni kao da su radili na istom sklopu utvrđenih problema i u skladu s istim sklopom utvrđenih kanona koje je najnovija revolucija u znanstvenoj teoriji i metodi učinila da izgledaju znanstvenima. (str. 146-147)

Gledana kroz prizmu Kuhnovog opisa udžbenika, kvantitativna metodologija i pripadajuća joj statistika u psihologiji predstavljaju jedinu zajedničku dodirnu točku svim psiholozima. Kvantitativne metode i statistika u psihologiji postaju psihologijski kanon iz kojeg se gradi čitava disciplina. Odnosno, kvantitativna metodologija je ahistorijski supstitut koji definira čitavu disciplinu te je ono što ju drži na okupu. Dewsbury (2009) to opisuje kada kaže:

U biti, većina psihologijskog rada odlikuje se onim što je Koch (1981) nazvao *ameaningful thought*. Taj pristup gleda na znanje kao na rezultat 'procesuiranja' a ne otkrića. Pretpostavlja da je znanje gotovo automatski rezultat majstorije, proizvodne trake, metodologije. (...) Nije slučajno da su na mnogim odjelima jedini zajednički osnovni kolegiji koje slušaju svi studenti u različitim područjima [psihologije] oni iz statistike i metodologije.

Ovakvo obrazovanje budućih stručnjaka i pisanje udžbenika je tipično za sve znanstvene discipline. To nije dovoljan razlog za kritiziranje korištenja kvantitativne metodologije i pripadajuće joj statistike u psihologiji. No, ovakav pogled na povijesni razvoj i ulogu kvantitativne metodologije stvara temelj za razmatranje jednog od statističkih postupaka koji ne samo da budi žučne polemike već desetljećima već, prema Cohenu (1994), usporava razvoj psihologije kao discipline, te ga čak aktivno nazaduje – taj postupak je testiranje značajnosti nul hipoteza (NHST)².

² Dalje u tekstu će se za izraz 'testiranje značajnosti nul hipoteza' koristiti akronim NHST koji je uvriježen u stranoj literaturi, a u originalu znači 'null hypothesis significance testing' (Nickerson, 2000).

Možemo ići toliko daleko i reći da je ahistorizacija metodološke povijesti psihologije i dovela do gotovo sveprisutnog mehaničkog korištenja testova značajnosti, bez dubljeg razmatranja njihove korisnosti ili značenja. Gigerenzer (2004, str. 590) ide toliko daleko da korištenje testova značajnosti naziva nul ritualom koji „ne postoji ni u prirodnim znanostima, ni u statistici“ već je njegov „izvor u umovima pisaca statističkih priručnika u psihologiji, obrazovanju i drugim društvenim znanostima.“ Takvo nekritičko korištenje testova značajnosti (kako zaključuju prije citirani autori), unatoč višedesetljetnoj sustavnoj kritici u literaturi, primjer je nekritične i ahistorične primjene kvantitativne metodologije. To je stvaran primjer Kuhnova ahistorizirana kanona psihologijske znanosti.

O problemima korištenja NHST-a se u internacionalnoj literaturi već raspravlja desetljećima (Nickerson, 2000), te ćemo pregled te rasprave ukratko prikazati u slijedećim poglavljima.

1.2 Pogrešno shvaćanje testova značajnosti u psihologiji

Nickerson (2000) u uvodu svog preglednog rada o kontroverzama vezanim uz NHST navodi kako se postupci testiranja nul hipoteze koje kritizira u psihologiji koriste već više od sedamdeset godina, a same kontroverze o ideji NHST-a se općenito mogu pratiti već više od dva stoljeća. S obzirom na obujam literature i raznolikost kritika NHST-a, pregled literature u ovom radu će pratiti Wagenmakersovu (2007) podjelu koja dijeli kritičare NHST-a u dvije skupine – jednu psihologijsku, a drugu statističku.

Rasprave među psiholozima se uglavnom svode na probleme interpretacije, dok se rasprave među statističarima svode na probleme „formalne konstrukcije“ samog testa. No, prije pregleda statističkih i psihologijskih kritika, potrebno je napraviti kratki povijesni pregled kako je NHST kao što ga danas poznajemo nastao.

1.2.1 Povijesni pregled razvoja NHST

Gill (1999) u svojem radu iznosi kratki opis povijesnog razvoja postupaka testiranja nul hipoteze kakve danas poznajemo i koristimo. Prvo što je važno spomenuti da su trenutni postupci „sinteza dvije jako utjecajne ali nekompatibilne škole u modernoj statistici“ (Gill,

1999, str. 649). Te dvije škole su ona Fishera, te Neymana i Pearsona. Prema Gillu (1999), Fisher je razvio postupak koji nam daje razine značajnosti na temelju podataka, dok su Neyman i Pearson osmislili rigidan proces za donošenje odluka koji ima za cilj potvrđivanje ili odbacivanje specifičnih a priori hipoteza.

Gigerenzer eksplicira točan postupak koji je Fisher predložio u svojim radovima (Fisher, 1955; Fisher, 1956; prema Gigerenzer, 2004, str. 590):

1. Odrediti statističku nul hipotezu. Nul hipoteza ne treba biti zaista nulta (nulta razlika).
2. Odrediti točan nivo značajnosti (npr. $p=0,051$ ili $p=0,049$). Ne smije se koristiti konvencionalni nivo značajnosti poput 5%, te se ne govori u prihvatanju ili odbacivanju hipoteza.
3. Koristiti ovu proceduru samo ako se zna malo o problemu koji se istražuje.

Gigerenzer nadalje jasno navodi da je Fisherov postupak dosta drugačiji od rituala testiranja nul hipoteza kakvog danas poznajemo. Razlikuje se jer „nema određenu alternativnu hipotezu...stoga koncepti poput statističke snage, pogreške tipa II, i teoretskih veličina efekta nemaju mjesta u Fisherovu postupku“ (Gigerenzer, 2004, str. 590).

S druge strane, Neymanov i Pearsonov postupak je kasniji razvoj koji je nastao iz njihove kritike Fisherovog postupka testiranja nul hipoteza; prvenstveno zato što Fisher u svojem postupku ne specificira alternativne hipoteze (Gigerenzer, 2004; Gill, 1999). Neyman i Pearson (1950; 1957; prema Gigerenzer, 2004) su predložili nešto što se naziva teorijom donošenja odluka. Postupak koji oni predlažu je slijedeći (Gigerenzer, 2004, str. 590-591):

1. Odrediti dvije statističke hipoteze, H_1 i H_2 , te odlučiti o α , β i veličini uzorka prije eksperimenta, temeljeći te odluke na subjektivnim *cost-benefit* razmatranjima. Ona određuju područje odbacivanja svake od hipoteza.
2. Ako su podaci u području odbacivanja H_1 , prihvatiti H_2 ; inače prihvatiti H_1 . Treba uzeti u obzir da prihvatanje hipoteze ne znači da vjerujemo u nju, nego samo da se ponašamo kao da je istinita.
3. Korisnost ovog postupka je ograničena na situacije gdje imamo disjunkciju između hipoteza (npr. ili je $\mu_1 = 6$ ili $\mu_2 = 10$ je istinito) te gdje se može odlučiti za smislene alfe i bete.

Neymanov i Pearsonov postupak se očito razlikuje od Fisherovog, jer jasno eksplicira alternativnu hipotezu te unaprijed određuje razine prihvatljive greške. U njihovom postupku možemo prepoznati neke od statistika koji se često danas koriste u testiranju nul hipoteza, poput pogreške tipa I, pogreške tipa II i statističke snage (definirane kao $1 - \beta$).

Kada se ova dva postupka testiranja hipoteza jasno opišu jedan uz drugoga, jasno je kako ono što Gigerenzer (2004; Gigerenzer, Krauss i Vitouch, 2004) naziva ritual nul hipoteza, odnosno NHST, hibrid ova dva pristupa. Gigerenzer (2004) i Gill (1999) jasno opisuju taj hibrid koji se danas koristi. Prvi korak je određivanje jedne statističke hipoteze (nul hipoteze), u skladu sa Fisherovim postupkom (osim što je nul hipoteza uvijek hipoteza o nultoj razlici, što nije nužno u Fisherovu postupku). Taj prvi korak nije u skladu s Neyman i Pearsonovim postupkom, jer ne određuje alternativnu statističku hipotezu, α , β ni veličinu uzorka. Drugi korak, donošenje da-ne odluke, je u skladu s Neyman i Pearsonovim postupkom, osim što razina značajnosti prema njima ne bi trebala biti određena konvencionalno već nakon razmišljanja i informiranog određivanja α , β i veličine uzorka. Hibrid u prvom koraku nalikuje Fisherovom postupku, u drugom Neyman i Pearsonovom, ali u stvari ne odgovara niti jednom.

Gigerenzer (2004, p. 588), po istom modelu kako su prije prikazani Fisherov te Neyman i Pearsonov postupak testiranja hipoteza, prikazuje i nul ritual, kako ga on naziva, te dodaje treći korak koji naziva „jedinstvenim u cijeloj statističkoj teoriji“:

1. Odrediti statističku nul hipotezu o nepostojanju razlike u prosjecima ili o nultoj korelaciji. Ne specificirati predviđanja istraživačkih hipoteza ili bilo koje druge alternativne supstantivne hipoteze.
2. Koristiti 5% kao konvenciju za odbacivanje nul hipoteze. Ako je značajno, prihvatiti istraživačku hipotezu. Prijaviti rezultate kao $p < 0,05$, $p < 0,01$, ili $p < 0,001$.
3. Uvijek provoditi ovaj postupak.

Zanimljivo je da niti Fisher ni Neyman i Pearson ne bi odobravali ovakvu sintezu (Gill, 1999, str. 653): „Fisher je prigovarao ideji određivanja razine značajnosti unaprijed te obaveznom procesu donošenja odluka s dva izlaza. Neyman i Pearson se nisu slagali sa interpretacijom p -vrijednosti (ili još gore, raspona p -vrijednosti određenih sa 'zvjezdicama') kao vjerojatnost pogreške tipa I jer ne predstavljaju dugoročnu vjerojatnost odbacivanja.“ Tako viđena, NHST koja se danas sveprisutno koristi je sinteza dva pristupa koja statističari

koji su ih osmislili ne bi odobrali, koja ima upitnu statističku osnovu te interpretacijsku vrijednost.

Hubbard i Ryan (2000) su povijesni razvoj i širenje korištenja testiranja nul hipoteza istražili studijom 12 psihologijskih časopisa od 1911. do 1988. U svojoj studiji dolaze do zaključka da „postotak empirijskih članka koji koriste NHST je u prosjeku 17% od 1911. do 1929, doseže 85% do 1960, i prelazi 90% sedamdesetih godina prošlog stoljeća“ (Hubbard i Ryan, 2000, str. 665). U drugoj studiji, Hubbard (2008) je procijenio da u razdoblju od 1990-2002 u 12 APA-inih časopisa, čak 94% empirijskih studija koristi testove statističke značajnosti. Studije poput ovih potvrđuju da je NHST, unatoč svojoj statističkoj upitnosti i interpretativnoj vrijednosti, kraljica statističkih analiza moderne psihologijske znanosti. Gigerenzer (2004, str. 589) to potvrđuje kada kaže da „šesnaesta edicija visoko utjecajnog udžbenika, Gerrigova i Zimbardova *Psychology and Life* (2002), prikazuje nul ritual kao vrhunsku statistiku te ga naziva 'kralježnicom psihologijskog istraživanja.'“

S obzirom na upitnost NHST kao sinteze postupaka Fishera i Neymana-Pearsona te njenu sveprisutnost, u slijedeća dva poglavlja je napravljen pregled kritika statističke osnove testiranja nul hipoteza, te pregled kritika njihove interpretacije. S obzirom da je fokus diplomskog rada na problemima interpretacije i poučavanja te interpretacije studentima, statistički problemi će se samo taksativno navesti uz kratka objašnjenja temeljena na radu Wagenmakersa (2007).

1.2.2 Kritika statističke osnove testova značajnosti u psihologiji

Kritike statističke osnove postupka testiranja nul hipoteze nastalog kao sinteza Fisherovog i Neyman-Pearsonovog pristupa su mnogostruke. Većina ih dolazi od istraživača koji zagovaraju bayesijanske postupke testiranja hipoteza (npr. Howard, Maxwell, i Flemming, 2000; Kruschke, 2010; Lecoutre, 1999; Wagenmakers, 2007).

Prikazani statistički problemi NHST-a su direktno citirani iz rada Wagenmakersa (2007, str. 782-788). Svaki problem je ukratko objašnjen, no izvodi formula, statistički dokazi i primjeri nisu prevođeni i citirani s obzirom da nisu uža tema ovog rada. Važno je napomenuti da su Wagenmakersove kritike prvenstveno upućene Fisherovom postupku, a ne kasnijoj sintezi Fisherova i Neyman-Pearsonovog postupka jer, prema Wagenmakersovim

riječima (2007, str. 781) „je [Fisherov postupak] filozofski bliže onome što eksperimentalni psiholozi vjeruju da rade kada analiziraju svoje podatke (odnosno kada istražuju svoje hipoteze).“ Očito Wagenmakers smatra da hibridni postupak koji se inače koristi niti ne zasluđuje ozbiljnu statističku kritiku.

Problem 1 – p -vrijednosti ovise o podacima koji nisu nikada izmjereni

P -vrijednost se ne računa samo na temelju statistika za opažene (izmjerene podatke), nego isto tako na hipotetičkim podacima koji su izvedeni, ali nikada zaista opaženi. Hipotetički podaci su oni koji se očekuju pod pretpostavkom H_0 , bez kojih bi bilo nemoguće konstruirati raspodjelu dobivenu uzorkovanjem (eng. *sampling distribution*). Ovo dovodi do donošenja različitih zaključaka temeljenih na istim podacima, ovisno koja raspodjela dobivena uzorkovanjem se koristi.

Problem 2 – p -vrijednosti ovise o mogućim ali nepoznatim subjektivnim namjerama istraživača

Ukratko, ovaj problem se svodi na to da p -vrijednost ovisi o planu uzorkovanja. Ako je plan uzorkovanja otvoren (npr. „u slučaju p -vrijednosti iznad 0,05, uključujem sve više i više sudionika u eksperiment dok ne postignem značajnost ispod 0,05 ili 0,01“), p -vrijednost ne ovisi samo o podacima ni o veličini efekta, nego i nacrtu istraživanja. Ako plan uzorkovanja nije definiran prije samog istraživanja i prikupljanja podataka, ne možemo odrediti točnu p -vrijednost.

Problem 3 – p -vrijednosti ne kvantificiraju statističke dokaze

Wagenmakers ne smatra da p -vrijednosti predstavljaju prihvatljivu kvantifikaciju većeg ili manjeg statističkog dokaza za hipotezu. Njegov primjer je slijedeća situacija: dvije studije imaju jednaku p -vrijednost (npr. $p=0,032$), ali jedna ima manji uzorak (npr. $n=12$) dok druga ima veći uzorak (npr. $n=98$). Koja od studija predstavlja snažniji statistički dokaz? Intuitivna logika većine psihologa je da je to ona koja ima veći uzorak. No, kada uzmemo u obzir da p -vrijednost ovisi o veličini efekta (što je veća veličina efekta, to je p niža) te veličini uzorka (što je uzorak veći, to je p niža); ne bi li studija sa manjim uzorkom trebala značiti veću veličinu efekta, pa tom logikom i snažniji statistički dokaz? Wagenmakers zastupa to stajalište, i kasnije ga detaljnije razrađuje kroz bayesijansku statistiku.

1.2.3 Kritika interpretacije testova značajnosti u psihologiji

Tema ovog rada je drugi smjer kritike NHST-a - analiza toga kako psiholozi interpretiraju testove značajnosti.

Jacob Cohen (1994) u svojem radu provokativnog naslova *Zemlja je okrugla* ($p < .05$) ukratko iznosi osnovni problem pri interpretaciji i shvaćanju NHST-a. Cohen (1994, str. 997) smatra da se ono što psiholozi očekuju od testa značajnosti drastično razlikuje od onog što zaista dobivaju: „Ono što želimo znati je: 'S obzirom na ove podatke, kolika je vjerojatnost da je H_0 istinita?' Ali ono što većina nas zna je da nam test zapravo govori: 'Pretpostavljajući da je H_0 istinita, koja je vjerojatnost ovih (ili još ekstremnijih) podataka?' Ove dvije stvari nisu iste...“

Iako se ova razlika može činiti kao semantička i sitničava, prema autorima koji kritiziraju NHST, ima velike implikacije za interpretaciju i korištenje testova značajnosti. Cohen nadalje objašnjava zašto je ovakvo shvaćanje NHST-a problematično kroz dva primjera silogizama, odnosno miješanja probabilističke logike NHST-a s modus tollensom. Primjeri koje Cohen nudi su slijedeći:

Ako je nul hipoteza istinita, onda se ovi podaci (D) ne mogu pojaviti.

Podaci su se pojavili.

Dakle, nul hipoteza je neistinita.

Ovo je valjani modus tollens, odnosno dokaz pomoću negacije konsekvensu u propozicijskoj logici. Kada bi ovo zaista bila logička forma testiranja nul hipoteza, ona ne bi bila problematična. No pri NHST, prva premisa je probabilističkog karaktera, što dovodi do nevaljanog zaključka, a ne valjanog modus tollensa. Ovo je logička forma pogrešne interpretacije NHST-a:

Ako je nul hipoteza istinita, pojavljivanje ovih podataka je malo vjerojatno.

Podaci su se pojavili.

Dakle, nul hipoteza je malo vjerojatna.

Pretvarajući prvu premisu u probabilističku informaciju, ovaj zaključak prestaje biti valjan. Cohen daje izvrstan svakidašnji primjer zašto je tomu tako bez ulaženja u formalne logičke izvode toga zašto je taj zaključak pogrešan. Za potrebe diplomskog rada, generiran je logički istovjetan primjer, samo bliži hrvatskom kontekstu:

Ako je osoba Hrvat, vjerojatno nije član Hrvatskog sabora.

Ova osoba je član Hrvatskog sabora.

Dakle, ova osoba vjerojatno nije Hrvat.

Prva premissa je istinita – većina Hrvata nisu članovi sabora. Dakle, na temelju istinite tvrdnje je izveden neistinit zaključak, da osoba koja je član Hrvatskog sabora vjerojatno nije Hrvat, iako svi članovi Hrvatskog sabora imaju hrvatsko državljanstvo. Silogizam sa saborom i silogizam o NHST prije njega su logički istovjetni, dakle, oboje su nevaljani zaključci jer negacija konsekvensa ne rezultira valjanim zaključkom ako je prva premissa probabilističkog karaktera.

Ovakav logički dokaz sustavnog problema s interpretacijom NHST-a je koristan, no potrebno je razmotriti i stvarne interpretacijske pogreške koje se događaju u istraživanjima. Nickerson (2000, str. 246) u svom radu opisuje niz takvih pogrešnih uvjerenja koje psiholozi imaju pri interpretaciji p -vrijednosti dobivenih testovima statističke značajnosti. Uvjerenje kojem posvećuje najviše pažnje te ga smatra najraširenijim je sljedeće:

Uvjerenje da p -vrijednost predstavlja vjerojatnost da je nul hipoteza istinita te da $1 - p$ predstavlja vjerojatnost da je alternativna hipoteza istinita.

S obzirom na prije opisane silogizme i opisivanje razlika između Fisherova i Neyman-Pearsonovog postupka, već je gotovo intuitivno jasno da nam NHST ništa ne govori o vjerojatnosti nul hipoteza, pa tako ni alternativnih hipoteza. Nickerson (2000, str. 248) tvrdi ne samo da nam NHST ne daje informacije o ovim vjerojatnostima, već se pri njegovom korištenju uvjeravamo kako dobivamo te odgovore dok sam test zaista odgovara na pitanje koje nitko uglavnom ni ne pita: „Koja je vjerojatnost dobivanja rezultata ekstremnog kao ovaj koji imamo u rezultatima ako pretpostavimo da je nul hipoteza istinita?“ Iz prije navedenog Gigerenzerova primjera Fisherovog postupka, u stvari dobivamo informaciju koja nam je minimalno vrijedna te ćemo se za nju interesirati tek ako ne znamo ništa o pojavi koju

ispitujemo, a to je da su ti podaci malo vjerojatni ako pretpostavimo da razlike (ili korelacije) ne postoje.

Kao jedan specifičan slučaj ovog uvjerenja, trebalo bi uključiti i interpretaciju da nam NHST govori kolika je vjerojatnost da su opaženi podaci dobiveni slučajno (Lambdin, 2012, str. 75-76). Ovo uvjerenje se isto temelji na ideju da testom statističke značajnosti implicitno dobivamo informacije o vjerojatnosti neke hipoteze (nul hipoteze; koristeći prije objašnjenju nevaljani zaključak koji miješa modus tollens s greškom uslijed korištenja premise s probabilističkim informacijama). Naime, podaci nam ne govore ništa o vjerojatnosti toga da su dobiveni po slučaju, jer se NHST po definiciji računa kao da je *nul hipoteza istinita* – podsjećamo na definiciju NHST, a to je da testira vjerojatnost dobivenih podataka *pod pretpostavkom da je nul hipoteza istinita*. Dakle, ako bismo sudili o vjerojatnosti toga jesu li podaci dobiveni po slučaju, sudili bismo o vjerojatnosti nul hipoteze koja je po definiciji jednaka 1 pri izračunu NHST-a.

Argument koji su iznijeli istraživači koji zagovaraju korištenje NHST kao metode za odbacivanje nul hipoteza unatoč prije iznesenoj kritici se najčešće naziva korelacijski argument (Trafimow i Rice, 2009). Zagovarači korelacijskog argumenta tvrde da iako nam p -vrijednost u formalno-statističkom smislu ne govori ništa o vjerojatnosti hipoteze, postoji visoka korelacija između visoke p -vrijednosti (dakle, male vjerojatnosti da se opaženi rezultati pojavljuju pod pretpostavkom da nul hipoteza vrijedi - $p(D|H_0)$) i male vjerojatnosti nul hipoteze ($p(H_0|D)$). Dakle, ako je statistik koji govori o podacima pod pretpostavkom nul hipoteze nizak, vjerojatno je i vjerojatnost nul hipoteze niska. Trafimow i Rice (2009) su napravili studiju u kojoj su računalno generirali 65.000 slučajnih setova podataka. Za svaki set podataka su izračunali vrijednosti $p(D|H_0)$ i $p(H_0|D)$ te izračunali korelaciju između njih, koja je iznosila $r=0,396$ ($r^2=15,6\%$). Kada su izračunali korelaciju samo unutar uvriježene razine značajnosti od 5%, ona je pala na $r=0,289$ ($r^2=8,4\%$). Njihova studija pokazuje da što je razina značajnosti niža, to je korelacija između $p(D|H_0)$ i $p(H_0|D)$ isto niža – te se postotak objašnjene varijance smanjuje. Ova studija upućuje na to da je korelacijski argument pogrešan, te pripisuje NHST-u neka svojstva koja ne posjeduje.

Uvjerenje da iz p -vrijednosti možemo zaključivati o replikabilnosti studije.

Drugo često pogrešno uvjerenje je ideja da p -vrijednost dobivena putem NHST nešto govori o replikabilnosti studije (Cohen, 1994; Gigerenzer, 2004; Nickerson, 2000; Schmidt, 1996). Oakes (1986) je ispitao 70 psihologa-istraživača i doktoranda znači li p -vrijednost od

0,01 da će se u budućim replikacijama u 99% slučajeva ostvariti isto značajan rezultat. S tom tvrdnjom se složilo 60% sudionika. Očito su zanemarili da je statistik koji nam zaista može dati informaciju o replikabilnosti studije u budućnosti je statistička snaga, a ne p -vrijednost (Schmidt, 1996). Ovakvo uvjerenje da nam p -vrijednosti i NHST govori o replikabilnosti studija je, uz manji prestiž koji se vezuje uz replikacijske studije (Lyndsay i Ehrenberg, 1993), mogući razlog zašto se takve studije u psihologiji uglavnom ne izvode. Studija Makela, Pluckera i Hegartyja (2012) je pokazala da u 100 psihologijskih časopisa od 1900-ih do 2010-ih, tek se 1,57% publikacija može smatrati replikacijama. Takvo ozračje dovodi do stanja koje opisuje Lambdin (2012, str. 77):

Da su urednici naših časopisa više naglašavali replikacije nego sumnjivo zahtijevali male p -vrijednosti, stvari bi danas možda izgledale drugačije. Nažalost, časopisi uglavnom žele „originalne“ ideje, ne replikacije, iako je vjerojatno da nemoguće očekivanje za „originalnošću“ svakog znanstvenika samo ohrabruje intelektualnu prevaru i razvodnjava sveukupnu kvalitetu objavljenih istraživanja (Andreski, 1972). Nadalje, dobra replikacija koja ne ostvari statističku značajnost se tako i tako neće objaviti, iako rezultati studije – dok god je metodologija u redu – ne bi trebali biti faktor koji utječe na to hoće li studija biti objavljena ili ne (Mahoney, 1977).

Kada se ova kriva interpretacija replikabilnosti vezuje uz NHST i posljedični nedostatak replikacija u psihologiji, te se uzme u obzir publikacijska pristranost (eng. *publication bias*) – tendencija da se samo objavljuju statistički značajni rezultati nauštrb onih neznačajnih (Ferguson i Heene, 2012) – postavlja se pitanje postoji li zaista empirijska provjera teorija koje trenutno postoje u psihologiji. Ferguson i Heene (2012, str. 555) idu toliko daleko da psihologiju nazivaju područjem „nemrtvih teorija“ (eng. *undead theories*) koje su „ideološki popularne ali nemaju svoju utemeljenost na činjenicama.“ Ova kritika se ne temelji na starim kritikama o manjoj vrijednosti društvene znanosti zbog njezinog subjekta istraživanja, s kojim se psihologija često (neosnovano) susretala kroz svoju povijest, nego na upitnim metodološkim i statističkim postupcima koji su se institucionalizirali u disciplini, nauštrb kvalitetnog znanstvenog rada i znanstvene kreativnosti.

1.3 U obranu NHST

Određen broj autora je stao u obranu korištenja NHST kao inferencijalnog postupka u psihologiji (npr. Chow, 1998; Hagen, 1997). Hagen u svom radu direktno odgovara prije citiranom Cohenovu (1994) članku, tako da ćemo zbog konzistentnosti rada predstaviti nekoliko njegovih argumenata.

Hagen (1998, str. 16) tako nudi kontra argument uvodeći u raspravu pitanje populacije i uzorka – on tvrdi kako hipoteze u NHST-u ne mogu govoriti o uzorku, već uvijek govore o populaciji: „Ako su nul hipoteza i alternativna hipoteza tvrdnje o uzorcima, određeni uzorak koji smo uzorkovali bi određivao prirodu našeg svijeta. Samo taj argument bi trebao biti dovoljan da nas odvraća od definiranja H_0 i H_1 u terminima uzorka. Ali uzmite u obzir što bi se dogodilo kada bismo uzorkovali dva uzorka. Mogli bismo završiti u svijetu u kojem i nul hipoteza i alternativna hipoteza pokušavaju biti istinite u isto vrijeme. Takav svijet je ne samo kapriciozan, nego nemoguć.“ Ovdje se već nazire fundamentalna razlika u shvaćanju toga što nul hipoteza (i alternativna hipoteza) zaista jesu – kritičari citirani u prijašnjim poglavljima ih promatraju iz perspektive 'koliko povjerenja mogu u njih imati?', dok ih Hagen postavlja u terminima 'ideje koje opisuju našu stvarnost'. S obzirom na to kako ih shvaćanju, gotovo potpuno suprotno interpretiraju i rezultate testova koje na njima primjenjuju.

Drugi Hagenov argument (1998, str. 17-19) upućuje baš na takvu fundamentalnu razliku u shvaćanju. On raspravu o NHST-u smješta kao diskusiju između dvije škole shvaćanja vjerojatnosti. Jedna, koju naziva fisherovska, vjerojatnost shvaća u terminima opaženih frekvencija. Druga, bayesijanska, vjerojatnost shvaća u terminima uvjerenosti u pojedinu hipotezu. S obzirom da kritike NHST-a opisane u prošlim poglavljima zaista uglavnom dolaze iz „kampa“ bayesianaca, ovo je logičan zaključak.

Treći Hagenov (1999; str. 21-22) argument koji ćemo spomenuti se odnosi na problem uvođenja probabilističkih sudova u modus tollens. Ono što Hagen naglašava jest da to što zaključak sa probabilističkim sudovima nema formalnu valjanost, ne znači da *ne može* biti istinit. Formalna logička valjanost znači da ako je argument predstavljen u konzistentnoj formi, on je uvijek valjan ali to ne znači da argumenti ne mogu biti „razumni i obranjivi“ ako nisu valjani u formalnom smislu. Dakle, Hagen se slaže da se NHST ne temelji na formalno valjanom modusu tollensu, ali to ne znači da zaključci koje donosimo na temelju njega nisu istiniti ili vrijedni. U svakodnevnom životu, iz kojeg Hagen i navodi primjere, ovo je svakako istinito.

Kako god, Hagenova obrana NHST-a zahtjeva svojevrstan komentar.

Najveći problem Hagenovog kontraargumenta je to što je on filozofski naivan. Odnosno, njegovo shvaćanje „objektivne“ stvarnosti – on koristi riječ svijet (ili populaciju) – je pomalo naivno i pojednostavljeno. Hagen polazi od idealne pretpostavke kako postoji populacijska vrijednost nezvezano od postupka uzorkovanja – pa tako ako hipoteza govori o

uzorku stvara više mogućih svjetova koji su međusobno nekompatibilni ili čak potpuno suprotni. Kao da je na trenutak zanemario da je znanost organizirani induktivni pothvat baš kako bi se dokučila ta stvarna populacijska vrijednost koja nam je nepoznata, i o njoj samo donosimo zaključke. Postuliranje 'nemogućih' svjetova i zaključivanje o njima na temelju uzoraka *je* sukus znanstvenog istraživanja. Hagen jednim potezom negira to da istraživač zaključuje o populaciji na temelju uzorka, pozivajući se na poprilično primitivna shvaćanja objektivne stvarnosti koja postoji nezavisno od načina na koji ju spoznajemo i promatramo. Takvo shvaćanje Warburton (1999, str. 105), s obzirom na modernu filozofiju znanosti, naziva i kritizira kao „jednostavno shvaćanje znanstvene metode“ koje je „začudo rasprostranjeno čak i među djelatnim znanstvenicima.“ Hagenova interpretacija NHST-a je naizgled direktno riješila problem indukcije, jer naime prema njemu, NHST govori direktno o populaciji, bez posredovanja uzorka – ili, što je mnogo vjerojatnije, predstavlja naivnu anakronu manifestaciju logičkog pozitivizma. I točno ovo uvriježeno shvaćanje testiranja nul hipoteze je interpretacijski problem koji možemo izvući iz bayesijanske kritike NHST-a, iako se možda ne slažemo sa njihovim statističkim ili praktičnim kritikama. Testovi značajnosti bi trebali stvarati više ili manje uvjerljive argumente o hipotezama, a ne se uzdizati kao neupitan objektivni put k spoznaji istine (u psihologijskom žargonu, populacijskih vrijednosti). O stvarnim vrijednostima ćemo uvijek zaključivati na temelju podataka, jer je to sama induktivna priroda znanosti, i svaka metoda (statistička ili neka druga) koja se poziva na populaciju na ovaj način mora biti dovedena u pitanje.

Manje filozofsko-znanstvenom terminologijom, u onom što Hagen naziva fisherovska škola shvaćanja vjerojatnosti postoji mnogo interpretacijskih nejasnoća koje su stvarni fokus ovog diplomskog rada. NHST, ako se razriješe problemi hibridizacije Fisherova i Neyman-Pearsonova postupka, ima unutarnju statističku logiku – ali ozbiljniji problemi su oni interpretacijski i zaključci koji se donose na temelju te logike. Interpretacija ne mora biti temeljena na bayesijanskim shvaćanjima vjerojatnosti, no mora biti ograničena i potkrijepljena velikim dijapazonom dokaza od kojih je p -vrijednost najslabiji.

Hagenova obrana NHST-a kao dobrog načina zaključivanja unatoč tome što nema valjanost formalne logike (modus tollens) upućuje točno na to. Zaključak temeljen na testu značajnosti nije formalno valjan (ako su premise istinite, i zaključak jest) već induktivni argument koji može potkrijepiti naše uvjerenje ili ne mora. Takvo shvaćanje ga moguće približava 'drugom kampu', ako ne već na osnovnoj razini interpretacije testa, onda na njegovoj ulozi u znanosti. Imajući u vidu radove poput prije opisane studije Trafimowa i

Ricea (2009) (korelacijski argument), potrebno je ozbiljno preispitati temelje uvriježenih interpretacija testova značajnosti.

Hagen (1999, str. 22) svoj rad zaključuje ovako: „Logika NHST-a je elegantna, izvanredno kreativna, i duboko ukorijenjena u naše metode statističkog zaključivanja. Malo je vjerojatno da ćemo se ikada uspjeti osloboditi te logike čak i ako jednog dana odlučimo to učini.“ Ovakav zaključak, nakon prije opisanih kritika, zaista upućuje na postojanje dva kampa u raspravi, koji imaju *uvjerenja* o najboljoj metoda. Jedan je sveprisutan u obrazovanju i *mainstream* literaturi psihologije, dok drugi (pokrenut bayesijanskom kritikom) predstavlja protutežu i kritiku uvriježenih istraživačkih praksi. Odgovor o tome što je najbolja praksa se vjerojatno ne može naći niti u jednom kampu, već je najbolji put pokušati naći sintezu kroz kritičnost pri provođenju istraživanja, ali posebice pri obrazovanju budućih stručnjaka.

1.4 Kako poučavati testiranje značajnosti hipoteza?

S obzirom na sve opisane kontroverze i interpretacijske probleme, određeni autori nude jednostavne savjete kako poboljšati poučavanje testiranja značajnosti hipoteza (Gigerenzer, Krauss i Vitouch, 2004; Haller i Krauss, 2002; Thompson, 1999). Njihovi prijedlozi su detaljnije razrađeni kasnije u raspravi, ali ovdje se taksativno spominju kao prijedlozi dobrih praksi:

1. Poučavati odvojeno Fisherov postupak i Neyman-Pearsonov te spomenuti fenomen njihove hibridizacije i pripadajuće probleme.
2. Paralelno uz NHST poučavati i Bayesovo pravilo za testiranje kondicionalnih vjerojatnosti – kako bi se postavio kontrast i jasna distinkcija između $p(D|H_0)$ i $p(H_0|D)$.
3. Naglašavati razliku između statističke značajnosti i praktične značajnosti.
4. Naglasiti ograničenja p -vrijednosti i potrebu za navođenjem drugih mjera poput veličine učinka, statističke snage i intervala pouzdanosti.

5. Jasno izložiti problem replikabilnosti ograničenja i opasnosti u donošenju zaključaka o replikabilnosti na temelju p -vrijednosti. Općenito, ohrabrivati replikacije, a ne kulturu zaključivanja o replikabilnosti na temelju statistički pokazatelja.
6. Više koristiti jednostavnije statističke opise poput deskriptivnih opisa podataka i korištenja grafičkih prikaza – kompleksne inferencijalne metode nisu same sebi svrha.

1.5 Poučavanje testiranja značajnosti nul hipoteza – što je sa studentima?

Teško je identificirati povijesni trenutak u kojem je nastao hibrid Fisherova postupka i Neyman-Pearsonovog postupka koji se danas koristi u psihologiji. No, u svakom slučaju se može pratiti njegovo etabliranje u obrazovanju budućih stručnjaka. Ako shvaćamo udžbenike u psihologiji kao kuhnovska sredstva ahistorizacije znanosti i indoktrinacije budućih stručnjaka, udžbenička prezentacija i poučavanje NHST-a je savršen praktičan primjer za to. Gigerenzer (2004, str. 596-597) tako na samo tri stranice Nunallyeva statističkog udžbenika iz 1975., *Introduction to Statistics for Psychology and Education*, pronalazi cijeli niz tvrdnji koje opisuju što značajan rezultat na razini od 5% u stvari znači:

- „vjerojatnost da je opažena razlika stvarna“
- „mala vjerojatnost da su se opaženi rezultati pojavili uslijed greške“
- „statističku pouzdanost...sa šansama od 95 do 100 da će se opažene razlike zadržati uslijed daljnjeg istraživanja“
- „opasnost prihvatanja statističkog rezultata kao stvarnog kada je u stvari opažen uslijed greške“
- stupanj u kojem se eksperimentalni nalaz može smatrati „ozbiljnim“
- stupanj „povjerenja u stvarnost nalaza“
- „istraživač može imati 95% povjerenja da se aritmetička sredina uzorka u stvari razlikuje od populacijske aritmetičke sredine“
- „ako je vjerojatnost niska, nul hipoteza je malo vjerojatna“
- „sve su ovo različiti načini da se kaže ista stvar“

Gliner, Leech i Morgan (2002) su napravili studiju u kojoj su detaljno proučili dvanaest udžbenika kako bi istražili kako oni tretiraju temu NHST-a. Šest udžbenika se koristilo na kolegijima koji poučavaju istraživačke metode u pedagogiji (eng. *educational*

science), a šest udžbenika na kolegijima koji poučavaju statistiku. Svaki udžbenik su proučila dva procjenjivača koji su ga ocijenili po unaprijed određenim kriterijima, među ostalim i na koji način i spominju li uopće kontraverze vezane uz NHST. Testiranje nul hipoteza nije na kritičan način bilo prikazano niti u jednom udžbeniku za istraživačke metode u pedagogiji, dok ih je tek dva od šest statistička udžbenika spomenulo. Statistički udžbenici koji su spomenuli raspravu i ograničenja testiranja nul hipoteza, to su učinili tek na trećoj razini (od pet mogućih, gdje je nulta razina potpuno zanemarivanje i nespominjanje) koju su autori unaprijed opisali kao: „Knjige tek par puta spominju ovaj problem, nemaju naslova [koji posebno naglašavaju pitanje]...te nemaju ništa u popisu ključnih riječi o ovom pitanju. Dakle, relativno izolirana spominjanja se mogu lako propustiti.“ (Gliner, Leech i Morgan, 2002, str. 88).

Ni udžbenici koji se koriste u obrazovanju psihologa u Hrvatskoj vjerojatno ne bi ništa bolje prošli na ovakvom testu. Petz (1997/2007, str. 137, kurziv u originalu) tako u svojem udžbeniku *Osnove statističke metode za nematematičare* opisuje testiranje nul hipoteze na ovaj način: „Upotrebom t-testa (ili kojeg drugog) mi provjeravamo postavljenu hipotezu: ako nam t-test pokaže da razlika među aritmetičkim sredinama nije statistički značajna, time smo *potvrdili* nul-hipotezu. Naprotiv, ako je razlika statistički značajna, *oborili* smo nul-hipotezu.“ Kada ovo usporedimo sa Cohenovim silogizmom sa Hrvatima i saborom, jasno je koja je pogreška u Petzovoj logici (viđena iz bayesijanskog kampa), odnosno logici u udžbenicima koji poučavaju studente statistici. Iako nismo napravili sustavnu analizu udžbenika koji se koriste na hrvatskim katedrama psihologije, više pretraživanja hrvatskih psihologijskih časopisa nije urodilo niti jednim radom koji bi se bavio ovim kontroverznim pitanjem. Iz perspektive hrvatske psihologije, kao da višedesetljetna rasprava i više desetina stranih autora i velikana psihologije koji „lome koplja“ oko testiranja značajnosti uopće ne postoje. Dobra pretpostavka je dakle, kako hrvatski studenti psihologije imaju točno onakvu perspektivu NHST koja se nudi u statističkim udžbenicima i propagira kao dobra istraživačka praksa – pomalo shizofreno stanje koje smo prije opisali u Nunallyevu udžbeniku bez dubljeg razmatranja smisla i uloge testova značajnosti.

S obzirom na nedostatak pokrivenosti ove teme u hrvatskoj literaturi, te primjer kao onaj iz Petzova udžbenika, stanje u obrazovanju budućih stručnjaka i znanstvenika u psihologiji ne može biti zadovoljavajuće. Situacija koju Schmidt (1996, str. 128) opisuje prije gotovo dvadeset godina se od riječi do riječi vjerojatno može primijeniti na hrvatsku psihologiju:

Naše mlade generacije nadolazećih znanstvenika se još uvijek odgajaju sa starim, diskreditiranim metodama koje se pouzdaju na statističko testiranje značajnosti. Kada poučavamo studente kako da analiziraju i interpretiraju podatke u pojedinim studijama, još uvijek ih učimo da primjene t testove, F testove, hi-kvadrata, i ANOVA-e. Učimo ih iste metode koje su u zadnjih 40 godina onemogućile pronalaženje pravog smisla u podacima i istraživačkoj literaturi te smo time usporili razvoj kumulativnog znanja u psihologiji i društvenim znanostima.

Uzimajući u obzir Schmidtovu pomalo apokaliptičnu najavu, ovim istraživanjem smo odlučili ispitati kako studenti psihologije u Hrvatskoj razumiju testiranje nul hipoteza, te je li to razumijevanje povezano sa sadržajima kojima ih se poučava na fakultetima. Preuzeli smo kratki test iz studija Oakesa (1986) i Hallera i Kraussa (2002) koji ispituju razumiju li studenti NHST s obzirom na prije opisane kontraverze. Uz test smo studente tražili informacije o ostvarenom uspjehu na studiju te pomoću cjelovitih popisa statističkih i metodoloških kolegija na njihovim fakultetima i ocjene iz tih kolegija. Takvo testiranje je provedeno na katedrama za psihologiju sveučilišta u Zagrebu (na Filozofskom fakultetu i Hrvatskim studijima), Zadru, Rijeci i Osijeku. Ovakvim ispitivanjem ćemo u jednom istraživanju ponuditi presjek rezultata obrazovnih praksi vezanih uz NHST u Hrvatskoj, te pozicionirati hrvatske studente psihologije s obzirom na studiju provedenu u Njemačkoj i SAD-u. Ukratko, ovo istraživanje predstavlja kritičko otvaranje teme statističkog testiranja hipoteza u hrvatskoj psihologiji.

2. CILJ ISTRAŽIVANJA I HIPOTEZE

Cilj istraživanja je deskriptivno opisati razumijevanje testiranja nul hipoteza kod hrvatskih studenata psihologije na nacionalnom uzorku. Uz to, cilj je ispitati postoji li povezanost između sadržaja koji se uče na statističkim i metodološkim kolegijima na fakultetu i razumijevanja testiranja nul hipoteza kod studenata, te postoji li povezanost između stava o testiranju nul hipoteza i razumijevanja testiranja nul hipoteza kod studenata.

Definirani su niže navedeni istraživački problemi:

1. Deskriptivno opisati razumijevanje testiranja nul hipoteza kod studenata psihologije u Hrvatskoj.

Hipoteza: Hrvatski studenti psihologije uglavnom ne razumiju testiranje nul hipoteza na razini koja bi im omogućila detektiranje ograničenja NHST-a opisanih u ovom radu.

2. Ispitati postoji li povezanost između razumijevanja testiranja nul hipoteza i općeg uspjeha na preddiplomskom studiju psihologije.

Hipoteza: Ne postoji povezanost između razumijevanja testiranja nul hipoteza i općeg uspjeha na preddiplomskom studiju psihologije.

3. Ispitati postoji li povezanost između razumijevanja testiranja nul hipoteza i prosječnog uspjeha na dosad odslušanim statističkim i metodološkim kolegijima na studiju psihologije.

Hipoteza: Ne postoji povezanost između razumijevanja testiranja nul hipoteza i prosječnog uspjeha na dosad odslušanim statističkim i metodološkim kolegijima na studiju psihologije.

4. Ispitati postoji li povezanost između razumijevanja testiranja nul hipoteza i stava prema testiranju nul hipoteza.

Hipoteza: Ne postoji povezanost između razumijevanja testiranja nul hipoteza i stava prema testiranju nul hipoteza.

5. Ispitati postoji li povezanost između razumijevanja testiranja nul hipoteza i glavnog interesa u psihologiji.

Hipoteza: Ne postoji povezanost između razumijevanja testiranja nul hipoteza i glavnog interesa u psihologiji. Čak ni studenti zainteresirani za znanstveno-istraživački rad ne razumiju testiranje nul hipoteza bolje od ostalih studenata.

3. METODA

3.1 Sudionici

U istraživanju je sudjelovalo 350 studenata psihologije sa pet fakulteta u Hrvatskoj – Hrvatskih studija Sveučilišta u Zagrebu (Hrvatski studiji), Filozofskog fakulteta Sveučilišta u Zagrebu (FFZG), Filozofskog fakulteta Sveučilišta u Rijeci (FFRI), Filozofskog fakulteta Sveučilišta u Osijeku (FFOS) i Odsjek za psihologiju Sveučilišta u Zadru (SuZD). Jedini hrvatski fakultet sa katedrom psihologije koji nije obuhvaćen uzorkom je onaj na Hrvatskom katoličkom sveučilištu. Studenti s tog fakulteta nisu uključeni u istraživanje jer u vrijeme provođenja istraživanja još niti jedna generacija nije imala dovoljno odslušanih statističkih i metodoloških kolegija.

Sudionici su u rasponu dobi od 20 do 28 godina, uz prosječnu dob od 22 godine ($M=22,15$, $SD=1,34$). S obzirom na rod, 89,7% studenata se identificiralo kao žensko ($n=314$) te manji dio kao muško ($n=33$) ili nekog drugog roda ($n=2$).

Zastupljenost studenata sa preddiplomskog (54%) i diplomskog studija (46%) je podjednaka. U tablici 1 se može vidjeti struktura uzorka s obzirom na fakultete te očekivana struktura, koja je napravljena na temelju broju ponuđenih mjesta za bruceše na pripadajućim katedrama psihologije za upis u akademsku godinu 2013./2014. (Postani student, 2013). Struktura u očekivanom i dobivenom uzorku je podjednaka, uz malo veću zastupljenost FFZG-a nauštrb studenata iz Rijeke, Osijeka i sa Hrvatskih studija. Mogući razlozi ove razlike su različite upisne kvote u prijašnjim godinama, fluktuacije broja studenata s obzirom na prolazak/pad godina i uvjete za upis na diplomatske studije te seljenju studenata sa jednog preddiplomskog studija na diplomski studij na drugom fakultetu.

Tablica 1. Sudionici istraživanja prema fakultetima

	Fakultet					
	Hrvatski studiji	FFZG	FFRI	FFOS	SuZD	Ukupno
Sudionici (n)	94	146	35	43	32	350
Sudionici (%)	26,9%	41,7%	10%	12,3%	9,1%	100%
Broj mjesta za bruceše (n)	50	83	37	30	35	235
Broj mjesta za bruceše (%)	21,3%	35,3%	15,7%	12,8%	14,9%	100%

3.2 Instrumenti³

3.2.1 Test pogrešnog shvaćanja testova statističke značajnosti (NHST test)

Test korišten za ispitivanje shvaćanja testiranja značajnosti hipoteza preuzet je iz rada Hallera i Kraussa (2002) i Oakesa (1986) te preveden na hrvatski za potrebe diplomskog rada. Ovaj test se često citira u radovima o NHST kontroverzama opisanim u uvodu (npr. Cohen, 1994; Gigerenzer, 2004; Gigerenzer, Krauss i Vitouch, 2004; Nickerson, 2000; Schmidt, 1996) te se njegovi rezultati smatraju kao indikativnim za razumijevanje NHST-a. NHST test se sastoji od uvodnog statičkog problema koji nalikuje na zadatak kakve studenti psihologije često susreću tijekom svojeg obrazovanja. Problem glasi ovako:

Pretpostavimo da istražuješ tretman za koji sumnjaš da ima utjecaj na izvedbu u nekom zadatku. Uspoređuješ aritmetičke sredine kontrolne skupine i eksperimentalne skupine (recimo da se svaka skupina sastoji od 20 sudionika). Nadalje, pretpostavimo da koristiš jednostavan t-test na nezavisnim uzorcima i dobiješ rezultat ($t = 2.7$, $df = 18$, $p = 0.01$). Molim označi niže navedene tvrdnje kao „točne“ ili „netočne“. „Netočno“ znači da se ta tvrdnja ne može logički izvesti iz prije navedenih premisa. Isto tako, uzmi u obzir da više njih ili niti jedna tvrdnja ne mora biti točna.

Nakon problema slijedi šest dihotomnih čestica koje sudionik može označiti kao 'točne' ili 'netočne', s obzirom na podatke u prije navedenom problemskom zadatku.

1. U potpunosti si pobila nul hipotezu (koja glasi: ne postoje razlike između populacijskih vrijednosti aritmetičkih sredina).	T	N
2. Odredila si vjerojatnost toga da je nul hipoteza točna.	T	N
3. U potpunosti si dokazala eksperimentalnu hipotezu (koja glasi: postoje razlika između populacijskih vrijednosti aritmetičkih sredina).	T	N
4. Iz prije navedenih podataka možeš izvesti vjerojatnost toga da je eksperimentalna hipoteza istinita.	T	N
5. Znaš koja je vjerojatnost donošenja pogrešne odluke u slučaju da odlučiš odbaciti nul hipotezu.	T	N
6. Posjeduješ pouzdan eksperimentalni nalazak u smislu da, hipotetski, ponoviš eksperiment mnogo puta u 99% slučajeva bi dobila značajan rezultat.	T	N

³ Sve formulacije u upitnicima u ovom istraživanju koje se odnose na rod su u ženskom rodu. S obzirom na rodnu strukturu populacije studenata psihologije (u uzorku gotovo 90% sudionica), ženski rod je odabran kao primjereniji. Pošto je takav izričaj suprotan uvriježenim normama, unaprijed je objašnjen svim sudionicima u uputi kako bi se izbjegao nesporazum. U Prilogu 1 je priložen cijeli upitnik.

Sve tvrdnje u upitniku su netočne. Jednostavna ispravna definicija NHST-a na kojoj Haller i Krauss odbacuju sve tvrdnje kao pogrešne jest: Test nam govori vjerojatnost opaženih podataka, pod pretpostavkom da je nul hipoteza istinita. Haller i Krauss (2002, str. 6), na temelju te definicije, daju kratka objašnjenja zašto je svaka od tvrdnji netočna (te su ova objašnjenja u skladu sa opisom krivih interpretacija NHST-a u uvodu ovog rada), osim objašnjenja za tvrdnju 6 koje smatra NHST verzijom Neyman-Pearsonovog postupka:

Tvrdnje 1, i 3

NHST nikada ne može dokazati ili ne dokazati hipoteze. On samo daje probabilističke informacije koje se mogu, u najboljem slučaju, interpretirati kao u skladu ili protivne teorijama i teorijskim očekivanjima. Kako god ih interpretirali, nikada na temelju njih ne možemo odbaciti hipotezu.

Tvrdnje 2, 4 i 5

Uglavnom je nemoguće dodijeliti neku vjerojatnost bilo kojoj hipotezi na temelju testova značajnosti. Ne možemo dati vjerojatnost 1 (tvrdnje 1 i 3) ili bilo koju drugu vjerojatnost (tvrdnje 2 i 4). Donošenje zaključaka o vjerojatnosti hipoteza je jedino moguće pomoću postupaka Bayesove statistike.

Tvrdnja 6

Tvrdnja 6 ispituje takozvanu „pogrešku replikacije.“ Gigerenzer (2004, str. 595) ovu tvrdnju objašnjava na slijedeći način: „Ovdje, $p=1\%$ insinuira da bi se takvi značajni podaci pojavili u 99% posto ponavljanja. To bi bilo istinito samo kada bi sa sigurnošću znali da je nul hipoteza istinita. U formalnim terminima, $p(D|H_0)$ se miješa sa $1 - p(D)$.“ Uglavnom, uvijek bi se trebalo izbjegavati korištenje testova značajnosti kao indikatora replikabilnosti rezultata, te koristiti druge mjere (primjer iz literature je statistička snaga) (Schmidt, 1996).

Unatoč tome što je test često citiran i korišten u literaturi, njegovi rezultati se u niti jednom radu do sada nisu interpretirali kao kompozitni rezultat na testu. U svim radovima su se odgovori sudionika tek deskriptivno analizirali za svaku pojedinu česticu. S obzirom na to, proveli smo faktorsku analizu kako bismo provjerili konvergiraju li čestice ka jednom faktoru – razumijevanju statističkih testova nul hipoteza. Matrica korelacija tvrdnji iz testa pokazuje niske korelacije između čestica kao što je vidljivo u tablici 2, što upućuje i na višefaktorsku strukturu te nisku unutarnju konzistentnost testa. Na isto upućuje i nizak Kaiser-Mayer-

Olkinov test prikladnosti matrice za faktorsku analizu (0,517). Važno je napomenuti da su korelacije dodatno snižene zbog smanjenog varijabiliteta dihotomnih čestica (kasnije je iz Tablice 3, vidljivo da su čestice uz to asimetrično raspodijeljene).

Faktorska struktura je nejasna – scree plot u Prilogu 2 upućuje na četverofaktorsku strukturu, dok Kaiser-Guttmanov kriterij i postoci objašnjene varijance upućuju na faktorsku strukturu od tri faktora ($\lambda_1=1,447$; $\lambda_2=1,115$; $\lambda_3=1,016$; $\lambda_4=0,975$). Rotirana trofaktorska struktura odgovara i grupiranju tvrdnji samih autora (Haller i Krauss, 2002) te drugih autora (Gigerenzer, 2004; Gigerenzer, Krauss i Vitouch, 2004) koji opisuju i grupiraju pogreške koje čestice testa mjere. Tako su se tvrdnja 1 i tvrdnja 3 grupirale na prvom faktoru; tvrdnja 2, tvrdnja 4 i tvrdnja 5 (moguće) na drugom faktoru, dok je tvrdnja 6 formirala zasebni četvrti faktor što je vidljivo iz tablice saturacija u Prilogu 2. Takva struktura je u skladu s pogreškama koje su opisali Haller i Krauss (2002). Provjereni su Cronbachovi alfe za tako formirane subskale, te su se oni pokazali u potpunosti nezadovoljavajući (za prvi faktor $\alpha=0,529$, te za drugi faktor $\alpha=0,237$).

Tablica 2. Matrica korelacija čestica NHST testa (n=334)

	Tvrdnja 1	Tvrdnja 2	Tvrdnja 3	Tvrdnja 4	Tvrdnja 5	Tvrdnja 6
Tvrdnja 1	1,000	-,089	,355	-,043	-,020	-,040
Tvrdnja 2	-,089	1,000	-,157	,168	,029	,016
Tvrdnja 3	,355	-,157	1,000	-,030	-,066	,074
Tvrdnja 4	-,043	,168	-,030	1,000	,034	-,008
Tvrdnja 5	-,020	,029	-,066	,034	1,000	-,001
Tvrdnja 6	-,040	,016	,074	-,008	-,001	1,000

S obzirom na nejasnu faktorsku strukturu te nezadovoljavajuće koeficijente pouzdanosti, nije bilo moguće raditi kompozitne rezultate na ovom testu, tako da je u daljnjim analizama svaka od šest čestica analizirana odvojeno.

Eliminacijska pitanja. U upitnik su uključena dva eliminacijska pitanja, kako bi se iz uzorka isključili sudionici koji nisu pohađali ili položili kolegije koji obuhvaćaju gradivo o testiranju nul hipoteza i drugim inferencijalnim statističkim postupcima. Na temelju eliminacijskih pitanja isključeno je 11 sudionika iz daljnjih analiza. Eliminacijska pitanja su glasila:

- a. Položila sam barem jedan statističko-metodološki kolegij koji obuhvaća inferencijalne statističke postupke.
- b. Tijekom studija susrela sam se s konceptom i barem jednim postupkom računanja nul hipoteze.

3.2.2 Stavovi prema testiranju nul hipoteza.

Za potrebe diplomskog rada konstruiran je kratak upitnik stavova prema statističkom testiranju nul hipoteza. Sastoji se od jednog zadatka rangiranja i šest čestica Likertovog tipa.

Zadatak rangiranja

Zadatak rangiranja je od sudionika tražio da procjene koji je prema njihovom mišljenju najvažniji statistički pokazatelj prema kojem procjenjuju je li neko istraživanje zaista dalo važan doprinos psihologijskoj literaturi. Ponuđena su im četiri statistička pokazatelja (veličina učinka, varijanca, test statističke značajnosti i veličina uzorka), te su ih trebali rangirati od 1 (najvažniji pokazatelj) do 4 (najmanje važan pokazatelj). Implicitna pretpostavka na kojoj je utemeljen ovaj zadatak je da što sudionik ima više povjerenja u testiranje nul hipoteza kao inferencijalno valjane metode, to će testove značajnosti procijeniti kao važnije u usporedbi s drugim statisticima.

Upitnik stava

Čestice Likertovog tipa u upitniku se procjenjuju od 1 do 5, gdje 1 označava potpuno neslaganje a 5 potpuno slaganje. Čestice su generirane kako bi obuhvatile gotovo idealiziran stav prema testiranju značajnosti nul hipoteza kao besprijekornom inferencijalnom postupku koje opisuje niz autora (Falk i Greenbaum, 1995; Gigerenzer, 1998; Gigerenzer, 2004; Lambdin, 2012; Monterde-i-Bort, Frías-Navarro i Pascual-Llobell, 2010) često ga nazivajući ritualnom primjenom statističkih postupaka (Gigerenzer, Krauss i Vitouch, 2004). Same čestice su navedene u tablici 3. Provjerena je homogenost skale te se pokazalo da upitnik ima nizak Cronbachov α ($\alpha=0,493$) te je stoga provedena faktorska analiza, kako bi se ispitalo konvergiraju li sve čestice jednom faktoru.

Preduvjeti za provedbu faktorske analize su ispunjeni – Bartlettov test sfericiteta je značajan (226,12 uz $df=15$ i $p<0,01$) a Kaiser-Meyer-Olkin iznosi 0,599. Provedena je analiza glavnih komponenata bez rotacije. Scree-plot je jasno upućivao na dvofaktorsku strukturu,

umjesto na očekivanu jednofaktorsku. Dvofaktorska struktura objasnila je 54,67% varijance šest čestica stava. Na temelju saturacija prikazanih u Tablici 3, vidljivo je da sve čestice osim druge i šeste konvergiraju ka prvom faktoru. Stoga su te dvije čestice izbačene iz kasnijih analiza.

Tablica 3. Deskriptivni podaci i faktorska opterećenja upitnika stavova prema testiranju nul hipoteza (n=339)

Čestica	M	SD	Faktor 1	Faktor 2
1. Postupak testiranja nul hipoteze najbolji je način statističkog zaključivanja u psihologijskim istraživanjima.	3,46	,970	0,746	-0,236
2. Mislim da studentski radovi iz psihologije koji dobivaju sveučilišne nagrade i priznanja (npr. rektorovu nagradu) moraju opovrgnuti nul hipotezu.	2,09	1,155	-0,004	0,800
3. Statističko testiranje nul hipoteze omogućuje napredovanje psihologije kao znanosti.	3,85	,803	0,717	-0,127
4. Testiranje nul hipoteze jedini je znanstveno opravdan način provjere istinitosti hipoteza u psihologiji.	2,50	1,139	0,671	0,104
5. Testiranje značajnosti nul hipoteza je prednost psihologije u odnosu na druge društvene znanosti.	3,34	1,021	0,616	0,170
6. Pri izradi diplomskog rada, bilo bi mi bitno opovrgnuti nul hipotezu.	3,02	1,274	0,122	0,787

Zadržane četiri čestice čine homogenu mjeru stava prema testiranju nul hipoteza. Cronbachov α je nešto viši u slučaju da se računa samo na četiri čestice prvog faktora ($\alpha=0,623$), što potvrđuje da je upitnik homogeniji bez dvije čestice koje ne konvergiraju prvom faktoru. Faktorska struktura zadržane četiri čestice jasno je jednofaktorska, a postotak objašnjene varijance četiriju čestica prvom glavnom komponentom iznosi 48,17% (scree plot prikazan u prilogu 3). Prosjek rezultata na četiri čestice se u kasnijim analizama koristi kao varijabla 'općeniti stav prema testiranju nul hipoteza.'

3.2.3 Upitnik demografskih i akademskih podataka

Za potrebe istraživanja je konstruiran kratki upitnik o osnovnim demografskim (dob, rod) i akademskim podacima. Akademski podaci su uključivali pitanja o: fakultetu, razini i godini studija, prosjeku ocjena na preddiplomskom studiju, glavnom interesu u psihologiji i ocjenama iz dosad odslušanih statističkih i metodoloških kolegija. Kod ispitivanja glavnog interesa u psihologiji, ponuđene su četiri opcije koje smo pretpostavili da su najčešće (klinička psihologija, organizacijska psihologija, znanstveno istraživački-rad te školska psihologija) te opciju 'drugo' kao pitanje otvorenog karaktera. Od sudionika se tražilo da odabere samo jednu opciju.

Za prikupljanje podataka o ocjenama iz dosad odslušanih metodoloških i statističkih kolegija napravljen je popis s imenima obaveznih kolegija koji obuhvaćaju ta područja za svaki od fakulteta. Popisi su formirani na temelju preddiplomskih i diplomskih programa studija psihologije objavljenih na stranicama pripadajućih odsjeka⁴.

Formirani popisi su dodatno dani na provjeru studentima ili zaposlenicima pripadajućih fakulteta, kako bi se provjerila njihova točnost. U slučaju da niti jedan sudionik nije odslušao i dobio ocjenu iz nekog od kolegija na popisu, taj kolegij je u potpunosti izbačen iz daljnjih analiza (npr. to je bio slučaj sa kolegijem *Multivarijatna statistika* na Filozofskom fakultetu u Rijeci).

Sudionici su morali upisati ocjenu koju su ostvarili iz tog kolegija. Uz to, studenti koji su pali od nekih kolegija, morali su upisati i ocjenu '1'. Na temelju upisanih ocjena '1' formirana je posebna varijabla koja je podijelila studente na one koji su pali barem jedan statistički ili metodološki kolegij, te na one koji su ih sve prošli. 294 studenata nije palo niti jedan kolegij, 36 je palo jedan, 7 je palo dva, jedan student je pao tri kolegija te isto tako jedan student je pao četiri kolegija. Ako su u međuvremenu položili kolegij, u računanju prosjeka se koristila prolazna ocjena.

⁴ Na stranicama odsjeka su pronađeni izvedbeni programi ili prihvaćeni prijedlozi programa (Studij psihologije, Hrvatski studiji Sveučilišta u Zagrebu, 2005; Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Zagrebu, 2012a; Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Zagrebu, 2012b; Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Rijeci, 2012a; Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Rijeci, 2012b; Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Osijeku, 2012a; Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Osijeku, 2012b; Odsjek za psihologiju Sveučilišta u Zadru, n. d.; Manenica i Sorić, 2005).

3.3 Postupak

Istraživanje je provedeno u ljetnom semestru akademske godine 2012/2013, tijekom travnja i svibnja. Prikupljanje podataka se provelo u četiri grada na pet fakulteta. Studentice i studenti su uzorkovani tijekom predavanja ili vježbi u sklopu obaveznih kolegija na samim fakultetima, na 3. godini preddiplomskog studija i 1. godini diplomskog studija. Testiranje je trajalo 10 do 15 minuta, uz kratku verbalnu uputu o cilju i prirodi istraživanja.

Zbog financijskih i vremenskih ograničenja, prikupljanje podataka su proveli asistenti i profesori na pripadajućim fakultetima (to je bio slučaj u Zadru, Osijeku, diplomskom studiju psihologije na Hrvatskim studijima, te sa jednom skupinom studenata na Filozofskom fakultetu u Zagrebu), zaposlenici sveučilišta (Rijeka) ili sam istraživač (preddiplomski studij na Hrvatskim studijima i Filozofski fakultet u Zagrebu). Svi upitnici su dostavljeni u Zagreb poštom te potom obrađeni.

4. REZULTATI

4.1 Uspješnost u studiju: prosjeci ocjena općih i metodoloških kolegija

U tablici 4 su prikazani prosjeci ocjena na preddiplomskim studijima, prosjeci ocjena iz metodoloških kolegija te postotak studenata na svakom fakultetu koji je pao barem jedan metodološki kolegij. Što se tiče prosjeka ocjena na preddiplomskim studijima, katedre na četiri fakulteta u Hrvatskoj su uglavnom ujednačene sa prosjekom bliskim 4, te jedino Hrvatski studiji odskoču sa nešto višim prosjekom od 4,365.

Kod prosjeka metodoloških predmeta, FFRI ima najniži prosjek od 2,916, dok najviši prosjek imaju opet Hrvatski studiji – 3,671. SuZD, FFOS i Hrvatski studiji tvore skupinu sa višim prosjecima, dok FFZG i FFRI imaju nešto niže metodološke prosjeke.

Uz to je izračunat i postotak studenata koji je pri ispunjavanju popisa s ocjenama naznačio da je barem pao jedan metodološki kolegij. SuZD ovdje odskoče od drugih fakulteta, sa izrazito visokim postotkom studenata koji su pali barem jedan kolegij. Ovo moguće i objašnjava mali broj studenata na svakoj godini na SuZD.

Tablica 4. Aritmetičke sredine prosjeka ocjena po fakultetima, prosjeka ocjena iz metodoloških kolegija i postotak sudionika koji su pali barem jedan kolegij

Fakultet	M	SD	N	M _(metod.)	SD _(metod.)	N _(metod.)	% padova
Hrvatski studiji	4,365	0,349	93	3,671	0,548	94	16%
FFZG	3,935	0,476	144	3,305	0,590	146	8,2%
FFRI	3,962	0,422	35	2,916	0,498	35	17,1%
FFOS	3,944	0,349	43	3,573	0,506	43	9,3%
SuZD	4,031	0,314	16	3,638	0,603	32	46,9%
Ukupno	4,065	0,454	331	3,428	0,606	350	14,9%

Ukupni prosjek iz metodoloških kolegija je očekivano manji od ukupnog prosjeka iz svih kolegija, npr. na FFRI-u ta razlika iznosi gotovo cijelu ocjenu ($M=3,962 > M=2,916$).

U prilogu 4 su prikazani i prosjeci ocjena po svakom pojedinom metodološkom kolegiju na svakom od pojedinih fakulteta. U tablicama P2, P3, P4, P5 i P6 su prikazani prosjeci ocjena po svakom statističko-metodološkom kolegiju na pojedinim fakultetima, te broj sudionika koji je naveo svoju ocjenu iz tog kolegija. Ponuđeni su i minimumi i maksimumima prosjeka kao moguća naznaka osobne jednadžbe profesora koji ocjenjuje. Eklatantan primjer za to je kolegij *Kvalitativne metode* na Hrvatskim studijima, koji nema uopće varijance te mu je maksimum jednak minimumu. Što su kolegiji sa viših godina, to manje sudionika ima ocjene iz njih. Npr. ocjene iz kolegija *Uvod u deskriptivni i inferencijalnu statistiku* ima 32 studenata, dok iz kolegija *Psihologijski praktikum* ima tek 12 studenata. Ovaj pad se događa jer uzorak čine studenti diplomskih i preddiplomskih studija, pa kolegije sa viših godina mlađi studenti nisu još odslušali. Uz to, ovi kolegiji često imaju vezane upise (uvjet za upis kolegija s više godine je položen kolegij s niže godine).

4.2 Interesi u psihologiji

Tijekom ispitivanja glavnog područja interesa u psihologiji, osim četiri koja su bila ponuđena, sudionici su izrazili još deset interesa, te je šest sudionika ostavilo praznu rubriku ili dopisalo opciju 'ne znam'. U kategoriju kliničke psihologije su spojene kategorije psihoterapije, zdravstvene psihologije i forenzičke psihologije. U kategoriju organizacijske su spojene kategorije marketinga, vojne i sportske psihologije. Tako su dobivene četiri kategorije prikazane u grafu 1 – u kategorije znanstveno-istraživačkog rada i školske psihologije nije spojena niti jedna druga kategorija.

Najviše studenata na svakom fakultetu je zainteresirano za kliničku psihologiju – to je najizraženije na SuZD (67,74%) i FFZG (59,57%). Organizacijska psihologija je drugi najčešći interes na svim fakultetima, osim na SuZD (9,68%), dok su školska psihologija i znanstveno-istraživački rad na trećem i četvrtom mjestu na svim fakultetima. Najmanje zainteresirani za školsku psihologiju su studenti na FFRI-u (3,23%), dok su istovremeno najzainteresiraniji za znanstveno-istraživački rad (19,35%).

4.3 Stavovi prema statističkom testiranju nul hipoteza

Stavovi prema statističkom testiranju nul hipoteza su ispitivani na dva načina: jednim zadatkom rangiranja i upitnikom stavova.

4.3.1 Zadatak rangiranja

Zadatak rangiranja je glasio:

Koji su, po tvom mišljenju, najvažniji statistički pokazatelji da je neko istraživanje dalo zaista vrijedan doprinos psihologijskoj literaturi: **(a) veličina učinka, (b) varijanca, (c) test statističke značajnosti ili (d) veličina uzorka.**

Graf 2. Procjena važnosti statističkih pokazatelja kvalitete istraživanja u rangovima (n=334)

Sudionici su svaki od statističkih pokazatelja rangirali na skali od 1 do 4, gdje je rang 1 predstavljao najvažniji statistički pokazatelj kvalitete istraživanja, dok je rang 4 predstavljao najmanje važan statistički pokazatelj istraživanja. Rezultati procjena za svaki rang su prikazani u grafu 2.

NHST je uvjerljivo procijenjen kao najvažniji statistički pokazatelj kvalitete istraživanja (čak 272 procjene koje ga smještaju na prvo i drugo mjesto). Drugi najvažniji je veličina učinka (192 procjene koje ju smještaju na prvo i drugo mjesto). Varijanca i veličina uzorka su procijenjeni manje važnima. Ovaj zadatak pokazuje jasnu tendenciju studenata da procjenjuju NHST i veličinu učinka kao najvažnijim pokazateljima prema kojima sude je li neko istraživanje važan doprinos literaturi ili nije.

4.3.2 Upitnik stavova

Rezultati na upitniku stava prema testiranju značajnosti nul hipoteza pokazuju da studenti nemaju pretjerano izražen stav ni u jednom smjeru. Uz teoretski raspon od 1 do 5, $M=3,28$ uz $Sd=0,680$. U grafu 4 se može vidjeti raspodjela rezultata na upitniku stava.

4.4 Razumijevanje testiranja nul hipoteza

U tablici 5 su opisane čestice iz NHST testa. Čestice su srednje težine ili teške ($M=p$ =indeks lakoće). Tvrdnja na koju su sudionici najčešće davali točan odgovor je prva, a čak je i u tom slučaju 44% studenata dalo krivi odgovor. Težina čestica se jasno vidi iz p i q vrijednosti – tvrdnja 3 je najdiskriminativnija, tvrdnja 1 je najlakša, a sve ostale tvrdnje su srednje težine ili teške. Tvrdnja 5 je uvjerljivo najteža jer ju je čak 88% studenata pogrešno procijenilo kao istinitu.

Pošto su tvrdnje dihotomnog karaktera, potrebno je uzeti u obzir i faktor pogađanja. Ako se studenti sa ovakvim sadržajem nisu susreli do sada, te s obzirom na ograničeno vrijeme rješavanja, vjerojatno je u određenom broju slučajeva došlo do pogađanja. Zbog toga, indeksi ne odražavaju samo upoznatost studenata s problemom već i slučajni faktor pogađanja odgovora. To zasigurno smanjuje i korelacije samih čestica sa drugim varijablama u istraživanju.

Tablica 5. Deskriptivni podaci čestica Testa pogrešnog shvaćanja testova statističke značajnosti – NHST testa ($n=336$)

Čestica	p	q	M	Varijanca	SD
1. U potpunosti si pobila nul hipotezu (koja glasi: ne postoje razlike između populacijskih vrijednosti aritmetičkih sredina).	0,56	0,44	0,56	0,25	0,50
2. Odredila si vjerojatnost toga da je nul hipoteza točna.	0,33	0,67	0,33	0,22	0,47
3. U potpunosti si dokazala eksperimentalnu hipotezu (koja glasi: postoje razlika između populacijskih vrijednosti aritmetičkih sredina).	0,50	0,50	0,50	0,25	0,50
4. Iz prije navedenih podataka možeš izvesti vjerojatnost toga da je eksperimentalna hipoteza istinita.	0,31	0,70	0,31	0,21	0,46
5. Znaš koja je vjerojatnost donošenja pogrešne odluke u slučaju da odlučiš odbaciti nul hipotezu.	0,12	0,88	0,12	0,11	0,33
6. Posjeduješ pouzdan eksperimentalni nalazak u smislu da, hipotetski, ponoviš eksperiment mnogo puta u 99% slučajeva bi dobila značajan rezultat.	0,42	0,58	0,42	0,24	0,49

Napomena: Točan odgovor na sve tvrdnje je odgovor 'netočno'.

Rezultati hrvatskih studenata psihologije su uspoređeni sa rezultatima studenata iz studije Hallera i Kraussa (2002) te je ta usporedba prikazana u grafu 5. Kao što je vidljivo iz grafa, hrvatski studenti su na svim tvrdnjama više griješili od njihovih njemačkih kolega. Najmanja razlika, od 10%, je kod najlakše prve tvrdnje dok je najveća razlika (30%) kod tvrdnje 3. Ovu usporedbu treba uzeti s oprezom, pošto hrvatski uzorak predstavlja nacionalni uzorak studenata psihologije, dok njemački uzorak predstavlja tek prigodno odabran manji broj grupa sa nekoliko njemačkih katedri psihologije.

Haller i Krauss (2002) u svojem radu spominju i proporciju studenata koji su odgovorili točno na svih šest tvrdnji. U njihovom uzorku, niti jedan student nije ispravno odgovorio na sve tvrdnje. U hrvatskom uzorku, to je uspjelo gotovo 1% studenata.

4.5 Povezanost razumijevanja testiranja nul hipoteza s interesima, stavovima i ocjenama

S obzirom da se NHST test pokazao nedovoljno konzistentnim za analize na ukupnom rezultatu, svaka čestica u njemu je analizirana pojedinačno s obzirom na povezanost s glavnim područjem interesa u psihologiji, stavovima prema statističkom testiranju nul hipoteza, padom barem jednog metodološkog kolegija te općim prosjecima ocjena i prosjecima ocjena iz metodoloških kolegija. Korištena je logistička regresija, te je za svaku česticu napravljen jedan model sa jednakim setom prediktora. Shema svih šest modela je slijedeća:

Kriterij: Čestica iz NHST testa (1-6)

Prediktor 1: Prosjek ocjena na preddiplomskom studiju

Prediktor 2: Prosjek ocjena na metodološkim kolegijima

Prediktor 3: Pad barem jednog metodološkog kolegija (dihotomna varijabla)

Prediktor 4: Općeniti stav prema testiranju nul hipoteza

Prediktor 5: Procjena važnosti statističkog pokazatelja kvalitete istraživanja – NHST (dihotomna varijabla)

Prediktor 6: Glavno područje interesa u psihologiji (dihotomna varijabla)

Varijabla glavnog područja interesa u psihologiji je kodirana kao dihotomna varijabla, gdje su sudionici grupirani u dvije skupine – onih sa interesom za znanstveno-istraživački rad i onih koji nisu izrazili takav interes.

Varijabla procjene važnosti statističkog pokazatelja kvalitete istraživanja je isto tako dihotomizirana – rang 1 i rang 2 su oformili skupinu visoke važnosti NHST pri procjeni kvalitete istraživanja, a rang 3 i rang četiri skupinu niske važnosti.

Varijabla pada barem jednog metodološkog kolegija je kodirano dihotomno – studenti koji nisu pali niti jedan kolegij su kodirani kao '0', a studenti koji su pali jedan ili više kolegija su kodirani kao '1'.

Kriterij je kodiran tako da veća vrijednost veća vrijednost označava odgovor 'netočno', a manja vrijednost odgovor 'točno'.

U tablicama 6, 7, 8, 9, 10, i 11 su prikazani rezultati logističkih regresija za svaku česticu. Model koji je objasnio najviše varijance je onaj za tvrdnju 2, gdje Naglekerkeov R^2 iznosi 5,4% varijance kriterija. U tom modelu se najpredikativnijom pokazala dihotomna varijabla pad/nema padova iz metodoloških kolegija ($\exp(B) = 3,084$). Taj prediktor je jedini značajan, no s obzirom na male postotke objašnjenja varijance, interpretacija ove statističke značajnosti kao praktične značajnosti bi bila baš ta pogrešna – fetišistička – interpretacija p -vrijednosti koja je tema ovog rada. S obzirom da je ukupni model objasnio manje od 6% varijance, opisivanje na razini pojedinih prediktora (pa čak i značajnih kao varijabla padova) nema nikakvu praktičnu vrijednost. Stoga, s obzirom na male postotke objašnjenja varijance, niti jedan prediktor ni model se ne može interpretirati kao dobar u opisivanju načina odgovaranja studenata na tvrdnje iz NHST testa.

Tablica 6. Regresijski model za tvrdnju 1: „U potpunosti si pobila nul hipotezu (koja glasi: ne postoje razlike između populacijskih vrijednosti aritmetičkih sredina).“ (n=306)

Prediktori	B	S.E.	Wald	df	Znač.	Exp(B)
Prosjek na preddiplomskom studiju	-,238	,385	,382	1	,537	,788
Prosjek na metodološkim kolegijima	,261	,301	,754	1	,385	1,299
Pad barem jednog metodološkog kolegija	-,515	,372	1,917	1	,166	,597
Općeniti stav prema testiranju nul hipoteza	-,329	,178	3,406	1	,065	,719
Procjena važnosti NHST	-,211	,318	,439	1	,508	,810
Glavni interes: znanstveno istraživački rad	,003	,346	,000	1	,992	1,003
Konstanta	1,694	1,247	1,847	1	,174	5,444
Cox & Snell R ²	0,027					
Nagelkerke R ²	0,036					
χ^2	8,384					
p	0,21					

Tablica 7. Regresijski model za tvrdnju 2: „Odredila si vjerojatnost toga da je nul hipoteza točna.“ (n=306)

Prediktori	B	S.E.	Wald	df	Znač.	Exp(B)
Prosjek na preddiplomskom studiju	,105	,409	,067	1	,796	1,111
Prosjek na metodološkim kolegijima	-,069	,316	,047	1	,828	,934
Pad barem jednog metodološkog kolegija	1,127	,375	9,048	1	,003	3,086
Općeniti stav prema testiranju nul hipoteza	,133	,185	,514	1	,474	1,142
Procjena važnosti NHST	,175	,333	,277	1	,599	1,192
Glavni interes: znanstveno istraživački rad	,193	,353	,299	1	,585	1,213
Konstanta	-1,590	1,312	1,469	1	,225	,204
Cox & Snell R ²	0,039					
Nagelkerke R ²	0,054					
χ^2	12,262					
p	0,56					

Tablica 8. Regresijski model za tvrdnju 3: „U potpunosti si dokazala eksperimentalnu hipotezu (koja glasi: postoje razlika između populacijskih vrijednosti aritmetičkih sredina).“ (n=305)

Prediktori	B	S.E.	Wald	df	Znač.	Exp(B)
Prosjek na preddiplomskom studiju	-,528	,396	1,777	1	,182	,590
Prosjek na metodološkim kolegijima	,565	,307	3,383	1	,066	1,759
Pad barem jednog metodološkog kolegija	-,411	,376	1,195	1	,274	,663
Općeniti stav prema testiranju nul hipoteza	-,239	,176	1,838	1	,175	,788
Procjena važnosti NHST	-,068	,313	,047	1	,828	,934
Glavni interes: znanstveno istraživački rad	,553	,351	2,479	1	,115	1,739
Konstanta	1,041	1,253	,691	1	,406	2,833
Cox & Snell R ²	0,039					
Nagelkerke R ²	0,053					
χ^2	12,253					
p	0,057					

Tablica 9. Regresijski model za tvrdnju 4: „Iz prije navedenih podataka možeš izvesti vjerojatnost toga da je eksperimentalna hipoteza istinita.“ (n=306)

Prediktori	B	S.E.	Wald	df	Znač.	Exp(B)
Prosjek na preddiplomskom studiju	,239	,412	,338	1	,561	1,271
Prosjek na metodološkim kolegijima	-,092	,318	,084	1	,772	,912
Pad barem jednog metodološkog kolegija	,519	,376	1,900	1	,168	1,680
Općeniti stav prema testiranju nul hipoteza	,000	,186	,000	1	,999	1,000
Procjena važnosti NHST	-,004	,329	,000	1	,991	,996
Glavni interes: znanstveno istraživački rad	,003	,360	,000	1	,994	1,003
Konstanta	-1,498	1,326	1,277	1	,259	,224
Cox & Snell R ²	0,08					
Nagelkerke R ²	0,011					
χ^2	2,386					
p	0,881					

Tablica 10. Regresijski model za tvrdnju 5: „Znaš koja je vjerojatnost donošenja pogrešne odluke u slučaju da odlučiš odbaciti nul hipotezu.“ (n=306)

Prediktori	B	S.E.	Wald	df	Znač.	Exp(B)
Prosjek na preddiplomskom studiju	,712	,613	1,349	1	,245	2,038
Prosjek na metodološkim kolegijima	-,634	,456	1,937	1	,164	,530
Pad barem jednog metodološkog kolegija	-,040	,564	,005	1	,943	,961
Općeniti stav prema testiranju nul hipoteza	-,155	,269	,331	1	,565	,857
Procjena važnosti NHST	,923	,629	2,152	1	,142	2,517
Glavni interes: znanstveno istraživački rad	-,193	,573	,114	1	,736	,824
Konstanta	-3,056	2,012	2,308	1	,129	,047
Cox & Snell R ²	0,015					
Nagelkerke R ²	0,030					
χ^2	4,766					
p	0,574					

Tablica 11. Regresijski model za tvrdnju 6: „Posjeduješ pouzdan eksperimentalni nalazak u smislu da, hipotetski, ponoviš eksperiment mnogo puta u 99% slučajeva bi dobila značajan rezultat.“ (n=305)

Prediktori	B	S.E.	Wald	df	Znač.	Exp(B)
Prosjek na preddiplomskom studiju	,351	,392	,803	1	,370	1,420
Prosjek na metodološkim kolegijima	-,053	,301	,031	1	,861	,949
Pad barem jednog metodološkog kolegija	,034	,370	,009	1	,926	1,035
Općeniti stav prema testiranju nul hipoteza	-,162	,175	,853	1	,356	,850
Procjena važnosti NHST	,093	,311	,090	1	,764	1,098
Glavni interes: znanstveno istraživački rad	,156	,338	,213	1	,645	1,169
Konstanta	-1,127	1,246	,818	1	,366	,324
Cox & Snell R ²	0,008					
Nagelkerke R ²	0,010					
χ^2	2,371					
p	0,883					

5. RASPRAVA

Glavni cilj ovog istraživanja je bio deskriptivno opisati razumijevanje statističkog testiranja nul hipoteza kod hrvatskih studenata na nacionalno reprezentativnom uzorku. To je postignuto primjenom NHST testa na pet katedri psihologije. Rezultati hrvatskih studenata psihologije su se na svih šest čestica testa pokazali lošijima od rezultata njihovih njemačkih kolega u istraživanju Hallera i Kraussa (2002). Zaključak Hallera i Kraussa (2002, str. 8) temeljen na nešto boljim rezultatima njemačkih studenata svakako onda još i više vrijedi za hrvatske studente psihologije: „...poučavanje studenata psihologije statistici na njemačkim sveučilištima ne uključuje učinkovito obrazovanje o *smislu* značajnosti.“ Gigerenzerovim (2004) riječima, primjena statističkih testova značajnosti nul hipoteza predstavlja „bezumnu statistiku.“ S obzirom da je ovo istraživanje provedeno na svim katedrama psihologije (osim novootvorene katedre na Hrvatskom katoličkom sveučilištu), loši rezultati studenata se ne mogu pripisati manje rigoroznom obrazovanju na pojedinim katedrama, niti nemaru njihovih profesora, već sustavnom zanemarivanju ove problematike kroz višedesetljetno pisanje udžbenika i poučavanje psihologa statistici. NHST je nekritički prihvaćen kao metoda koja vodi objektivnoj spoznaji, bez razmatranja toga kako je zaista nastala i za što je bila namijenjena. Tako viđena, NHST je ahistorijski⁵ mehanizam koji omogućuje napredak znanosti (navodno nam govori koje su hipoteze istinite a koje nisu), što s obzirom na kontroverze opisane u ovom radu, predstavlja veliki rizik za legitimnost psihologije kao discipline.

Ocjene, kako opći prosjeci tako i metodološki, se nisu pokazale prediktivnim za razumijevanja smisla NHST-a. Haller i Krauss (2002) i Gigerenzer (2004) nude shvaćanje zašto je tomu tako - kriva shvaćanja testova značajnosti studenti nasljeđuju od svojih profesora, te ih tako perpetuiraju u obrazovanju i znanstveno-istraživačkom radu. Haller i Krauss (2004, str. 8) na temelju svojeg istraživanja tako zaključuju da na njemačkim sveučilištima: „...četiri od pet profesora metodologije dijele kriva uvjerenje o ovom konceptu [NHST] sa svojim studentima.“ Prema viđenju Hallera i Kraussa, za ova kriva uvjerenja postoje jasni krivci a to su profesori psihologije koji poučavaju metodologiju. Ja bih se više priklonio konstruktivnijoj kritici Gigerenzera koji interpretira ovaj problem na razini sistemske manjkavosti u obrazovanju koja perzistira desetljećima. Gigerenzer to opisuje (2004, str. 596): „Ako su studenti 'naslijedili' iluzije od svojih učitelja, gdje su ih učitelji

⁵ Ahistorijski u kuhnovskom smislu, bez razmatranja kako je nastao ni čemu je služio u kontekstu kada je osmišljen.

pronašli? Odgovor je svima očigledan: U prvim udžbenicima koji su psihologe upoznali sa testiranjem nul hipoteza prije više od 50 godina.“ Ukratko, ocjene predstavljaju kvantitativnu procjenu uspješnosti savladavanja nastavnih sadržaja na pripadajućim kolegijima, ali ti sadržaji uopće ne promatraju NHST kao metodu ograničene primjene. NHST je ugrađena u metodološki korpus tehnika koje psiholozi koriste u svojim istraživanjima, te se njegovo korištenje etabliralo kao jedna od osnovnih inferencijalnih metoda za donošenje zaključaka.

Dodatna potvrda za sistematičnost poučavanja pogrešne interpretacije testova značajnosti su stavovi i interesi studenata psihologije. Pošto se NHST sustavno poučava kao provjereni inferencijalni mehanizam – Dewsbury (2009) takvu metodologiju u psihologiji naziva „proizvodnom trakom“ – studenti neće imati kritički izgrađene stavove prema istom. Stav se neće polarizirati na one koji su upoznati s tematikom i visoko su kritični te one koji su 'krivo uvjereni', već studenti uglavnom neće promišljati ovo pitanje kao nešto prema čemu bi uopće trebali imati stav. Ovo pitanje se u okviru kolegija ne raspravlja niti nudi kao kontroverzna tema o kojoj postoji čitava literatura objavljena kroz više desetljeća, već jednostavno kao metoda koja se koristi kako bi se donio sud – je li hipoteza istinita ili nije. Stoga, manjak prediktivne valjanosti stava predstavlja još jedan indikator manjka pokrivenosti ove teme u programima obrazovanja psihologa. Možemo ići toliko daleko i reći da je indikator za nepostojanje ove teme.

Zadatak procjene važnosti statističkih pokazatelja kao kriterija kvalitete istraživanja upućuje baš na to – studenti NHST uvjerljivo procjenjuju kao najvažniji statistički pokazatelj o kvaliteti istraživanja. Dakle, studenti psihologije kvalitetu nekog istraživačkog rada procjenjuju po rezultatu statističkog testa značajnosti – unatoč tome što nam p-vrijednost tog testa ništa ne govori o statističkog snazi, veličini efekta niti replikabilnosti tog rezultata.

Krajnji argument koji upućuje na ograničenost hrvatskog metodološkog obrazovanja za psihologe je i manjak prediktivnosti interesa u psihologiji za razumijevanja NHST-a. Ako pretpostavimo da se studenti koji su zainteresirani za znanstveno-istraživački rad više angažiraju oko metodoloških i statističkih kolegija i učenja tih sadržaja, nepostojanje razlike između njihovog razumijevanja NHST-a i ostalih studenata upućuje da sadržaja kritičnih prema NHST nema dovoljno ili nema uopće. Studenti se uglavnom ne upućuju na dublje razmatranje postupka testiranja značajnosti i shvaćanje njegova smisla. Svakako, NHST se poučava na razini vještine (znanja kako provesti test računski ili kako koristiti softver koji ga može provesti), ali razumijevanje na smislenoj razini nije prioritet.

Potrebno je naglasiti da ovakvi rezultati ne stavljaju Hrvatsku izvan praksi u drugim državama. Sistematični manjak refleksije o smislu testiranja nul hipoteza nije problem hrvatske psihologije, nego sveukupne psihologije kao discipline; pa tako i one u Hrvatskoj. Na taj zaključak upućuju slični rezultati dobiveni i u drugim studijama koji su koristile NHST test (Falk i Greenbaum, 1995; Haller i Krauss, 2002; Oakes, 1986). Hrvatska psihologija, nažalost, prati ove trendove u internacionalnom smislu, te su rezultati ovog istraživanja u skladu sa rezultatima u literaturi što se tiče studentske (ne)uspješnosti rješavanja NHST testa.

Pri budućim istraživanjima ove teme, ono što je uputno je razvijanje većeg i psihometrijski valjanog testa razumijevanja smisla testiranja značajnosti nul hipoteza. Prema literaturi (Nickerson, 2000) ali i prema interpretacijama samog Hallera i Kraussa (2002), u pozadini krivih interpretacija NHST-a se skriva više krivih uvjerenja o tome što nam p-vrijednosti zaista govore. Test koji bi specifično testirao svako od tih krivih uvjerenja bi jasnije razlučivao što stoji u samom korijenu krivih interpretacija NHST-a. Na temelju istih krivih uvjerenja, bilo bi dobro razviti i prateći upitnik stavova prema NHST. Uz to, umjesto uzimanja ocjene kojom se indirektno mjeri jesu li studenti uopće savladali neke sadržaje vezane uz kritičko promatranje interpretacije NHST-a, bilo bi zanimljivo sustavno ispitati nastavne sadržaje i udžbenike (npr. kao u studiji Gliner, Leech i Morgan, 2002). Tako bi se moglo, za eksperimentalni nacrt, oformiti dvije skupine studenata – jednu koju se poučava inferencijalnim postupcima uz kritični pregled testiranja NHST i savjetima dobrih praksi, te drugu koju se poučava na dosad uvriježeni način. Bilo bi zanimljivo ispitati razlike u shvaćanju NHST-a nakon takve intervencije.

Ono što je svakako svojevrstan zaključak ovog istraživanja jest da se tretman testiranja značajnosti nul hipoteza u obrazovanju psihologa treba revidirati. Haller i Krauss (2002, str. 10-15) nude relativno jednostavan pristup za objašnjavanje testova značajnosti studentima. Prvi korak je da se studente jasno poučava kako postoje dvije paradigme kako testirati hipoteze i podatke – NHST i bayesijanska statistika. U drugom koraku se studentima jednostavno objašnjava osnovu NHST-a – to da testira odgovaranje podataka (D) hipotezi (H). Izraženo kao kondicionalna vjerojatnost, ovo glasi $p(D | H_0)$. Jednostavna rečenica koja je jasna i točna interpretacija ove kondicionalne vjerojatnosti jest: vjerojatnost pojavljivanja opaženih podataka, pod pretpostavkom da je H_0 istinita. Ono što Haller i Krauss ne spominju, a bilo bi dobro uključiti u ovaj dio, je i razlika između Fisherova postupka i Neyman-Pearsonovog. Treći korak je poučavanje Bayesovog pravila, kojim se može testirati odgovara li hipoteza (H) opaženim podacima (D). Poučavanjem Bayesova pravila uz NHST se jasno

prikazuje da je njihova logika obrnuta – jedan testira odgovaranje podataka hipotezi, dok drugi testira odgovaranje hipoteze podacima.

No, kritiziranje i odbacivanje interpretacije NHST-a kao tehnike koja nam govori o istinitosti hipoteza se gubi puno jednostavnosti i 'elegancije' u testiranju podataka dobivenih psihologijskom istraživanjima. Postoji li neka jednako jednostavna i primjenjiva alternativa? Gigerenzer (2004, str. 604) je predložio dobar odgovor točno na to pitanje:

Znamo ali često zaboravljamo da problem induktivnog zaključivanja nema jedno rješenje. Nema uniformno najjačeg testa, odnosno, metode koja je najbolja za svaki problem. Statistička teorija nam daje kutiju alata s učinkovitim instrumentima, koji zahtijevaju razmišljanje o tome kada bi ih trebali koristiti. Kada udžbenici i kurikulumi počnu poučavati korištenju te kutije alata, studenti će automatski naučiti razmišljati. I shvatit će da u mnogim primjenama, vješta i transparentna deskriptivna analiza je dovoljna, i preporučljivija od primjena statičkih postupaka koji su odabrani zbog njihove kompleksnosti i teške dokučljivosti. Razmišljanje je dio umjetnosti statistike.

Thompson (1999) daje neke savjete o tome koje statističke metode koristiti uz ili umjesto NHST pri analizi podataka. U skladu s Gigerenzerovom procjenom, ne postoji jedna metoda koja će zamijeniti pogrešno interpretiran NHST. Ono što Thompson preporuča je korištenje veličina učinka uz p -vrijednosti testova značajnosti. Prema Thompsonu (1999, str. 170) veličine učinaka pomažu u „donošenju suda o praktičnoj i sadržajnoj značajnosti rezultata.“ Nadalje, kaže kako nam vjerojatnost događaja ništa ne govori o njegovoj važnosti jer „malo vjerojatni događaji nisu nužno i važni.“ Sustavno prijavljivanje veličina efekata, uz to, olakšava kasnije meta-analitičke studije literature koje ih mogu agregirati i koristiti u donošenju zaključaka.

Drugi savjet koji Thompson predlaže je veći fokus na replikacije. No, ne daje jasnu ideju kako bi ovo proveo, samo dijagnosticira problem (Thompson, 1999, str. 175): „Disciplina [psihologija] još nije riješila sve probleme vezane uz odlučivanje o dovoljnim zahtjevima za replikabilnost rezultata, opet moguće, zbog toga što su neki autori netočno pretpostavili da statistički testovi procjenjuju populaciju.“ Makel, Plucker i Hegarty (2012, str. 541) isto tako oklijevaju postaviti jasne statističke kriterija koji će upućivati istraživače na to je li potrebna replikacije ili nije – jer kako kažu, „[takve preporuke] se često krivo koriste i interpretiraju kao zakoni a ne oprezni prijedlozi.“ Nadalje, oni ne preporučaju „replikaciju svakog preddiplomskog završnog rada“ već „studija koje utječu na teoriju, važne politike, i/ili velike grupe ljudi kako bi dobili korisne i provokativne spoznaje“ (Makel, Plucker i Hegarty, 2012, str. 541).

Treći Thompsonov (1999) savjet je korištenje intervala pouzdanosti. Iako se nekritično korištenje intervala pouzdanosti može činiti sličnim nekritičnom korištenju NHST, prema Thompsonu (1999, str. 175), oni imaju jednu jasnu prednost a to je da bi se pri usporedbi više studija koje istražuju isti konstrukt sličnim metodologijama, iz preklapanja većeg broja intervala pouzdanosti mogu donijeti zaključci o pravim vrijednostima statističkih parametara.

Gigerenzer, Krauss i Vitouch (2004) imaju niz prijedloga za poboljšanje stanja u psihologijskoj literaturi. Prvo je testiranje većeg broja alternativnih hipoteza, i odlučivanje koja je najbolja između njih. Kao primjer ovakve prakse, spominju istraživanja Rieskamp i Hoffragea koji u svom radu testiraju osam hipoteza o tome kako ljudi predviđaju profite kompanija, te istraživanje Gigerenzera i Hoffragea u kojem testiraju predviđanja šest kognitivnih strategija u rješavanju problema (Gigerenzer, Krauss i Vitouch, 2004, str.14). Njihov drugi savjet je veće fokusiranje na smanjivanje nesustavnog varijabiliteta te nude niz strategija kako to postići pažljivijom izradom nacrtu istraživanja. Isto tako upućuju na korištenje jednostavnijih, deskriptivnih analiza podataka i grafova koje ponekad mogu biti jednako informativne kao kompleksnije metode poput analize varijance.

Gigerenzer, Krauss i Vitouch (2004, str. 20) u svom radu nude izvrstan završni komentar na polemiku o NHST-u i metodama koje bi se uz nju trebale koristiti:

Što bi se dogodilo da nema nul rituala ili NHST-a? Ništa se ne bi izgubilo, osim zbunjenost, anksioznosti i platforme za lijeno teoretsko razmišljanje. Mnogo bi se moglo dobiti, poput znanja o raznim statističkim oruđima, trening u statističkom razmišljanju, i motivaciju da se zaključuje o preciznim predikcijama vlastitih hipoteza. Bismo li trebali zabraniti nul ritual? Naravno – to je pitanje intelektualnog integriteta. Svaki istraživač bi trebao imati hrabrosti ne predati se ritualu, i svako urednik, pisac udžbenika, i savjetnik bi se trebali osjećati pozvano da promoviraju statističko razmišljanje i odbacuju bezumne rituale.

Preporuke na temelju ovog rada ne bi išle toliko daleko, ali se autori slažu u tome da bi se interpretacije NHST-a trebale pažljivo razmotriti i ubuduće poučavati iz kritičke perspektive. U svakom slučaju, potrebno je izbjeći ostvarenje Skinnerove optužbe upućene Fisheru kako „...poučavaju statistiku umjesto znanstvene metode“ (Skinner, 1972, str. 319).

6. ZAKLJUČAK

Glavni cilj ovog istraživanja bio je deskriptivno opisati koliko hrvatski studenti psihologije razumiju statističko testiranje nul hipoteza. Primjenom NHST testa se pokazalo da test ne ispunjava osnovne metrijske karakteristike, te da razumijevanje testiranja nul hipoteza ne predstavlja jednodimenzionalan konstrukt. Uputno bi bilo razviti duži test sa boljim metrijskim karakteristikama, koji će uključivati subskale za svako od krivih uvjerenja. Zaključak koji možemo izvući iz testa je da su rezultati hrvatskih studenata sustavno lošiji od rezultata u studiji Hallera i Kraussa (2002). Uz to, pokazalo se da opći uspjeh studenata te uspjeh u metodološkim i statističkim kolegijima ne predviđa koliko studenti razumiju statističko testiranje nul hipoteza. Stavovi studenata prema statističkom testiranju nul hipoteza, mišljenje o važnosti takvog testiranja te glavno područje interesa u psihologiji isto tako ne predviđaju uspješnost rješavanja NHST testa. Otvoreno je pitanje je li ovaj manjak prediktivnosti rezultat uslijed prevelikog pogađanja točnih odgovora, loših metrijskih karakteristika testa ili stvarna manjak kritičkog promišljanja testiranja nul hipoteza. Na temelju svega prije navedenog, općeniti zaključak jest da ova tema nije dovoljno obuhvaćena niti jednim programom psihologije u Hrvatskoj, te kako bi se njezinom poučavanju trebalo posvetiti više pažnje u okviru statističkog i metodološkog obrazovanja budućih psihologa. Studenti bi trebali biti obrazovani u kritične korisnike testova značajnosti, koji postavljaju jasna interpretacijska ograničenja pri njihovom korištenju, te bi njihovi profesori trebali imati testove na raspolaganju koji to mogu jednoznačno ispitati.

7. LITERATURA

- Chow, S. L. (1998). Precis of statistical significance: Rationale , validity , and utility. *Behavioral and Brain Sciences* 21, 169–239.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 997–1003.
- Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge, UK: Cambridge University Press.
- Dewsbury, D. A. (2009). Is Psychology Losing Its Foundations? *Review of General Psychology*, 13, 281–289.
- Falk, R., i Greenbaum, C. W. (1995). Significance Tests Die Hard: The Amazing Persistence of a Probabilistic Misconception. *Theory & Psychology*, 5(1), 75–98.
- Feyerabend, P. K. (1975/1993). *Against Method: Outline of an Anarchistic Theory of Knowledge*. London: Verso.
- Flis, I. (2012). U obranu psihologije: znanost kao žrtva svoje metode. *Diskrepancija*, 11(16/17), 67-82.
- Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33(5), 587–606.
- Gigerenzer, G. (1998). We need statistical thinking, not statistical rituals. *Behavioral and Brain Sciences*, 21(2), 199–200.
- Gigerenzer, G., Krauss, S. i Vitouch, O. (2004). The Null Ritual: What You Always Wanted to Know ABOUT Significance Testing but Were Afraid to Ask. U D. Kaplan (Ur.), *The Sage handbook of quantitative methodology for the social sciences* (str. 391-408). Thousand Oaks, CA: Sage.
- Gliner, J. A., Leech, N. L., i Morgan, G. A. (2002). Problems With Null Hypothesis Significance Testing (NHST): What Do the Textbooks Say? *The Journal of Experimental Education*, 71(1), 83–92.
- Gill, J. (1999). The insignificance of null hypothesis significance testing. *Political Research Quarterly*, 52(3), 647-674.

- Hagen, R. L. (1997). In praise of the null hypothesis statistical test. *American Psychologist*, 52(1), 15–24.
- Haller, H., i Krauss, S. (2002). Misinterpretations of Significance : A Problem Students Share with Their Teachers. *Science Education*, 7(1), 1–20.
- Howard, G. S., Maxwell, S. E., i Fleming, K. J. (2000). The proof of the pudding: an illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. *Psychological Methods*, 5(3), 315.
- Hubbard, R. (2004). Alphabet soup: Blurring the distinctions between p's and α 's in psychological research. *Theory & Psychology*, 14, 295–327.
- Hubbard, R., i Ryan, P. A. (2000). Statistical Significance with Comments by Editors of Marketing Journals: The Historical Growth of Statistical Significance Testing in Psychology and its Future Prospects. *Educational and Psychological Measurement*, 60(5), 661–681.
- Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, 14(7), 293-300.
- Kuhn, T. S. (1962/2002). *Struktura znanstvenih revolucija*. Zagreb: Naklada Jesenski i Turk.
- Lambdin, C. (2012). Significance tests as sorcery: Science is empirical - significance tests are not. *Theory & Psychology*, 22(1), 67–90.
- Lecoutre, B. (1999). Beyond the significance test controversy: Prime time for Bayes. *Bulletin of the International Statistical Institute: Proceedings of the Fifty-second Session of the International Statistical Institute*, 205-208.
- Machamer, P. (2002). A Brief Historical Introduction to the Philosophy of Science. U Machamer, P., i Silberstein, M. (Ur.), *The Blackwell Guide to the Philosophy of Science* (str. 1-17). Malden, MA: Blackwell Publishers Inc.
- Manenica, I., i Sorić, I. (2005). Program diplomskog studija psihologije. Pronađeno 3. 4. na: <http://www.unizd.hr/Portals/12/pdf/Psihologija-diplomski.pdf>

- Michell, J. (2003). The Quantitative Imperative Positivism, Naive Realism and the Place of Qualitative Methods in Psychology. *Theory & Psychology*, 13(1), 5-31.
- Milas, G. (2005). *Istraživačke metode u psihologiji i drugim društvenim znanostima*. Jastrebarsko: Naklada Slap.
- Monterde-i-Bort, H., Frías-Navarro, D., & Pascual-Llobell, J. (2010). Uses and abuses of statistical significance tests and other statistical resources: a comparative study. *European Journal of Psychology of Education*, 25(4), 429–447.
- Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological Methods*, 5(2), 241.
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. Chichester: Wiley.
- Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Osijeku (2012a). Izvedbeni nastavni plan za akademsku godinu 2012./2013 – preddiplomski studij. Pronađeno 3. 4. na: http://web.ffos.hr/psihologija/dat/s_111/File/PSIHOLOGIJA-PREDDIPLOMSKI-2012-2013.pdf
- Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Osijeku (2012b). Izvedbeni nastavni plan za akademsku godinu 2012./2013 – diplomski studij. Pronađeno 3. 4. na: http://web.ffos.hr/psihologija/dat/s_105/File/PSIHOLOGIJA-DIPLOMSKI-2012-2013.pdf
- Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Rijeci (2012a). Plan i program sveučilišnog preddiplomskog studija psihologije. Pronađeno 3. 4. na: <http://www.ffri.uniri.hr/wp-content/uploads/2013/02/PSIH-Preddiplomski-studij-2012.pdf>
- Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Rijeci (2012b). Plan i program sveučilišnog diplomskog studija psihologije. Pronađeno 3. 4. na: <http://www.ffri.uniri.hr/wp-content/uploads/2013/02/PSIH-Diplomski-studij-2012.pdf>
- Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Zadru (n.d.). Program preddiplomskog studija društvenih znanosti – smjer Psihologija. Pronađeno 3. 4. na:

<http://www.unizd.hr/psihologija/Studijskiprogrami/Preddiplomski/tabid/721/Default.aspx>

Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Zagrebu (2012a). Preddiplomski studijski program iz psihologije. Pronađeno 3. 4. 2013 na: <http://psihologija.ffzg.unizg.hr/uploads/57/63/5763q7gT23Y2SyBnpRFgPg/Program-preddiplomskog-studija-psihologije.pdf>

Odsjek za psihologiju, Filozofski fakultet Sveučilišta u Zagrebu (2012b). Diplomski studijski program iz psihologije. Pronađeno 3. 4. 2013 na: http://psihologija.ffzg.unizg.hr/uploads/y6/OU/y6OUUnQgVMPTg68VzlZrVIw/Program-diplomskog-studija-psihologije_v01102012_2.pdf

Petz, B. (1997/2007). *Osnove statističke metode za nematematičare*. Jastrebarsko: Naklada Slap.

Postani student (2013). Pronađeno 15. 2. 2013. na: <https://www.postani-student.hr/Ucilista/Default.aspx>

Skinner, B. F. (1972). *Cumulative record*. New York: Appleton-Century-Crofts.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72-101.

Studij psihologije, Hrvatski studiji Sveučilišta u Zagrebu (2005). Prijedlog Programa Preddiplomskog i Diplomskog studija psihologije. Pronađeno 3. 4. 2013 na: http://www.hrstud.unizg.hr/dokumenti/public/nastava/nastavni_program/Psihologija.pdf

Thompson, B. (1999). If Statistical Significance Tests are Broken/Misused, What Practices Should Supplement or Replace Them? *Theory & Psychology*, 9(2), 165–181.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, 14(5), 779–804.

Warburton, N. (1999). *Filozofija: Osnove*. Zagreb: Kruzak.

PRILOG 1 – UPITNIK (PRIMJER ZA HRVATSKE STUDIJE)

Pred tobom se nalazi kratak upitnik o stavovima i razumijevanju statističkog testiranja nul hipoteza. Ovaj upitnik se primjenjuje u okviru istraživanja za diplomski rad iz psihologije na Hrvatskim studijima Sveučilišta u Zagrebu. Diplomski rad izrađuje Ivan Flis, pod mentorstvom dr. sc. Ive Šverko. Podaci prikupljeni ovim upitnikom koristit će se jedino u svrhu istraživanja te ti je pri ispunjavanju zajamčena potpuna anonimnost.

Sve formulacije koje se odnose na rod u ovom upitniku su u ženskom rodu. Ženski rod je odabran, suprotno uvriježenoj konvenciji u hrvatskom jeziku, s obzirom da psihologiju uglavnom studiraju studentice; te ovim putem upućujem kolege da odgovaraju kao da su sve formulacije napisane i u muškom rodu.

Unaprijed te molim da se suzdržiš od prepisivanja i konzultiranja s kolegicama i kolegama tijekom ispunjavanja upitnika jer, kao i svaka studentica ili student psihologije, dobro znaš kako to može drastično narušiti niz metrijskih karakteristika ovog upitnika i same rezultate istraživanja. Unaprijed hvala na razumijevanju i sudjelovanju!

SVEUČILIŠTE I FAKULTET: Hrvatski studiji Sveučilišta u Zagrebu

DOB: _____

ROD: žensko ☐ muško ☐ drugo ☐

RAZINA STUDIJA preddiplomski ☐ diplomski ☐

GODINA STUDIJA _____

PROSJEK OCJENA NA PREDDIPLOMSKOM STUDIJU _____

GLAVNI INTERES U PSIHOLOGIJI (zaokruži jedan):

a. Klinička psihologija	d. Znanstveno-istraživački rad
b. Organizacijska psihologija	e. Školska psihologija
c. Drugo: _____	

Na linije pokraj imena kolegija upiši ocjenu koju si ostvarila ako si odslušala taj kolegij. Ako si neke od navedenih kolegija pala i ponovno ih slušaš tijekom ovog semestra, na pripadajuću liniju upiši ocjenu '1'. Ako si kolegij ponovno odslušala i položila, upiši ocjenu jedan i prolaznu ocjenu (npr. 1, 3). U slučaju da neke od navedenih kolegija još nisi uopće odslušala, ostavi pripadajuće crte prazne:

Uvod u znanstveno istraživački rad : _____

Praktikum iz istraživačkih metoda : _____

Mjerenje u psihologiji : _____

Faktorska analiza : _____

Eksperimentalne metode : _____

Teorija psihologijskih testova : _____

Neeksperimentalne metode : _____

Regresijska analiza : _____

Kvalitativne metode : _____

Izrada psihologijskog istraživanja : _____

Deskriptivna statistika : _____

Inferencijalna statistika : _____

Modeli analize varijance (ANOVA) : _____

- a. Položila sam barem jedan statističko-metodološki kolegij koji obuhvaća inferencijalne statističke postupke (npr. t-test, ANOVA-u ili hi-kvadrat). **DA / NE**
- b. Tijekom studija susrela sam se s konceptom i barem jednim postupkom računanja nul hipoteze. **DA / NE**

Sljedeća pitanja se odnose na tvoj stav prema statističkom testiranju nul hipoteze:

Koji su, po tvom mišljenju, najvažniji statistički pokazatelji da je neko istraživanje dalo zaista vrijedan doprinos psihologijskoj literaturi: **(a) veličina učinka, (b) varijanca, (c) test statističke značajnosti ili (d) veličina uzorka.**

U donjoj tablici rangiraj ove pokazatelje prema njihovoj važnosti (dovoljno je da uz svaki rang navedeš pripadajuće slovo pokazatelja):

Rang 1 (najvažniji statistički pokazatelj kvalitete istraživanja)	Rang 2	Rang 3	Rang 4 (najmanje važan statistički pokazatelj kvalitete istraživanja)

Pažljivo pročitaj tvrdnje i ocijeni svoj stupanj slaganja sa njima. Pri tome koristi sljedeću skalu za odgovore:

U potpunosti se **NE** slažem 1 2 3 4 5 U potpunosti se slažem

Postupak testiranja nul hipoteze najbolji je način statističkog zaključivanja u psihologijskim istraživanjima.	1 2 3 4 5
Mislim da studentski radovi iz psihologije koji dobivaju sveučilišne nagrade i priznanja (npr. rektorovu nagradu) moraju opovrgnuti nul hipotezu.	1 2 3 4 5
Statističko testiranje nul hipoteze omogućuje napredovanje psihologije kao znanosti.	1 2 3 4 5
Testiranje nul hipoteze jedini je znanstveno opravdan način provjere istinitosti hipoteza u psihologiji.	1 2 3 4 5
Testiranje značajnosti nul hipoteza je prednost psihologije u odnosu na druge društvene znanosti.	1 2 3 4 5
Pri izradi diplomskog rada, bilo bi mi bitno opovrgnuti nul hipotezu.	1 2 3 4 5

U ovom zadatku ti je ponuđen statistički problem kakvi su vrlo česti na studiju psihologije, posebice na metodološkim i statističkim kolegijima i vježbama. Pažljivo pročitaj zadatak i samostalno odgovori na pitanja.

Pretpostavimo da istražuješ tretman za koji sumnjaš da ima utjecaj na izvedbu u nekom zadatku. Uspoređuješ aritmetičke sredine kontrolne skupine i eksperimentalne skupine (recimo da se svaka skupina sastoji od 20 sudionika). Nadalje, pretpostavimo da koristiš jednostavan t-test na nezavisnim uzorcima i dobiješ rezultat ($t = 2.7$, $df = 18$, $p = 0.01$). Molim označi niže navedene tvrdnje kao „točne“ ili „netočne“. „Netočno“ znači da se ta tvrdnja ne može logički izvesti iz prije navedenih premisa. Isto tako, uzmi u obzir da više njih ili niti jedna tvrdnja ne mora biti točna.

7. U potpunosti si pobila nul hipotezu (koja glasi: ne postoje razlike između populacijskih vrijednosti aritmetičkih sredina).	T	N
--	---	---

8. Odredila si vjerojatnost toga da je nul hipoteza točna.	T	N
--	---	---

9. U potpunosti si dokazala eksperimentalnu hipotezu (koja glasi: postoje razlika između populacijskih vrijednosti aritmetičkih sredina).	T	N
---	---	---

10. Iz prije navedenih podataka možeš izvesti vjerojatnost toga da je eksperimentalna hipoteza istinita.	T	N
--	---	---

11. Znaš koja je vjerojatnost donošenja pogrešne odluke u slučaju da odlučiš odbaciti nul hipotezu.	T	N
---	---	---

12. Posjeduješ pouzdan eksperimentalni nalazak u smislu da, hipotetski, ponoviš eksperiment mnogo puta u 99% slučajeva bi dobila značajan rezultat.	T	N
---	---	---

Hvala ti na sudjelovanju u istraživanju! U slučaju da te zanimaju dobiveni rezultati, sama tema istraživanja ili koji su točni odgovori na pitanja u prošlom zadatku, slobodno se javi na **ivan.flis@student.hrstud.hr**

PRILOG 2

Graf P1. Scree plot za NHST testa

Tablica P1. Saturacije čestica NHST testa

	Faktor		
	1	2	3
Tvrdnja 1	,810	-,026	-,125
Tvrdnja 2	-,203	,656	,023
Tvrdnja 3	,812	-,077	,154
Tvrdnja 4	,109	,774	-,127
Tvrdnja 5	-,023	,388	,108
Tvrdnja 6	,023	,054	,975

PRILOG 3

Graf P2. Scree plot upitnika stavova prema testiranju nul (četiri čestice)

PRILOG 4

Kolegiji sa deset ili manje sudionika s ocjenama su izbačeni iz prikaza.

Tablica P2. Prosjeci ocjena na statističkim i metodološkim kolegijima – Hrvatski studiji

Kolegiji	N	Min	Max	M	Sd
Uvod u znanstveno istraživački rad	93	2	5	3,72	,799
Mjerenje u psihologiji	93	1	5	4,30	,734
Eksperimentalne metode	86	2	5	3,67	,804
Neeksperimentalne metode	76	2	5	3,91	,836
Kvalitativne metode	42	5	5	5,00	,000
Deskriptivna statistika	93	1	5	2,83	,916
Inferencijalna statistika	91	1	5	3,20	,922
Modeli analize varijance - ANOVA	87	1	5	3,60	,723
Praktikum iz istraživačkih metoda	41	2	5	3,63	,829
Regresijska analiza	24	1	5	3,58	1,283

Tablica P3. Prosjeci ocjena na statističkim i metodološkim kolegijima – FFZG

Kolegiji	N	Min	Max	M	Sd
Uvod u metodologiju eksperimentalne psihologije	145	2	5	3,49	,914
Neeksperimentalna psihologijska metodologija	69	1	5	4,10	,789
Statistika u psihologiji I	146	1	5	2,88	,862
Statistika u psihologiji II	145	1	5	2,90	,839
Psihologijski praktikum I	145	1	5	3,48	,980
Psihologijski praktikum II	143	1	5	3,57	,892
Uvod u teoriju testova	131	1	5	3,71	,940
Kvantitativna interpretacija testova	108	1	5	2,84	,787
Psihometrija	36	1	5	3,17	,910

Tablica P4. Prosjeci ocjena na statističkim i metodološkim kolegijima – FFRI

Kolegiji	N	Min	Max	M	Sd
Uvod u psihologijsku statistiku	35	1	5	2,57	,948
Metodologija psihologijskih istraživanja	35	2	5	3,43	,815
Psihologijska statistika	35	1	4	2,17	,747
Praktikum iz eksperimentalne psihologije 1	34	2	4	2,59	,701
Teorije mjerenja	35	2	5	3,94	1,136
Praktikum iz eksperimentalne psihologije 2	33	2	5	2,52	,795

Tablica P5. Prosjeci ocjena na statističkim i metodološkim kolegijima – FFOS

Kolegiji	N	Min	Max	M	Sd
Metodologija eksperimentalne psihologije	43	2	5	3,70	,887
Uvod u psihologijsku statistiku	43	2	4	2,53	,631
Inferencijalna statistika	43	1	4	2,56	,700
Praktikum iz eksperimentalne psihologije I	43	2	5	3,70	,773
Praktikum iz eksperimentalne psihologije II	43	2	5	3,91	,750
Mjerenje u psihologiji	43	2	5	4,21	,861
Teorija testova	43	1	5	3,84	1,090
Konstrukcija i interpretacija testova	24	4	5	4,63	,495
Metodologija eksperimentalne psihologije	43	2	5	3,70	,887

Tablica P6. Prosjeci ocjena na statističkim i metodološkim kolegijima – SuZD

Kolegiji	N	Min	Max	M	Sd
Osnove psihologijske metodologije	31	1	5	2,23	1,283
Uvod u deskriptivnu i inferencijalnu statistiku	32	1	5	3,31	,821
Praktikum eksperimentalne psihologije I	31	1	5	3,35	1,018
Osnove inferencijalne statistike	31	1	5	3,26	1,125
Praktikum eksperimentalne psihologije II	31	2	5	3,68	,945
Praktikum eksperimentalne psihologije III	20	2	5	4,00	,973
Teorije mjerenja	20	3	5	4,50	,607
Psihologijski praktikum	12	3	5	4,75	,622