**Interpretiranje rezultatov študentskih anket**

***Vasja Vehovar***

**Povzetek:** Podan je pregled statističnih in tudi nekaterih vsebinskih izzivov interpretacije rezultatov študentskih anket, ki so večinoma na petstopenjskih bipolarnih ordinalnih lestvicah. Pri tem se uporablja rezultate UL 2015/2016. Po predstavitvi vprašanj so najprej obravnavani mejniki pri interpretaciji povprečij, kjer je podana utemeljitev, zakaj so sta 3.5 in 4.0 zelo primerna mejnika, pomembni pa sta tudi meji 3.0 in 4.5. Nato je podan še pregled ostalih ključnih statistično-metodoloških vidikov: slučajne napake, sistematične napake, intervali zaupanja ter odnos med statistično značilnostjo in relativno pomembnostjo. Sledi pregled možnih vsebinskih dejavnikov, ki lahko vplivajo na ocene: pedagoška prizadevnost ter zunanji dejavniki, od všečnost pedagoga, do posebnosti predmeta, programa, študentov itd. V zadnjem delu so predstavljene nekatere primerjave, ki se najpogosteje izvajajo na rezultatih študentskih anket: primerjava komponent, primerjav predmetov istega pedagoga, primerjava med predmeti programa, primerjava med leti in primerjava med programi oziroma fakultetami. V tem okviru je podan sistematičen pregled statistično-metodoloških in vsebinskih izzivov interpretacij.

Vsebina

[1. Uvod 3](#_Toc532026402)

[2. Statistično metodološki vidiki interpretacij 3](#_Toc532026403)

[2.1 Predstavitev vprašanj in v študentskih anketah UL 3](#_Toc532026404)

[2.2 Mejniki pri interpretaciji povprečjij 5](#_Toc532026405)

[2.3 Slučajna napaka in statistična značilnost 8](#_Toc532026406)

[2.4 Sistematična napaka zaradi neodgovorov 10](#_Toc532026407)

[2.5 Relativna pomembnost razlik 11](#_Toc532026408)

[3. Vsebinski vidiki interpretacije 13](#_Toc532026409)

[3.1 Pedagoška prizadevnost 13](#_Toc532026410)

[3.2 Vsebinske interpretacije, dejavniki in problem kavzalnosti 14](#_Toc532026411)

[3.3 Pregled interpretacij dejavnikov za tipične primerjave 16](#_Toc532026412)

[A. Primerjave komponent znotraj izvedbe določenega predmeta 17](#_Toc532026413)

[B. Primerjave ocen za različne predmete istega pedagoga pri istih študentih v istem letu 18](#_Toc532026414)

[C0. Isti predmet, ista izvedba, istega pedagoga pri različnih študentih v istem letu 19](#_Toc532026415)

[C1. Isti predmet, različna izvedba, istega pedagoga pri različnih študentih v istem letu 19](#_Toc532026416)

[C2. Različni predmeti istega pedagoga pri različnih študentih v istem letu 20](#_Toc532026417)

[D0. Različni predmeti in različni pedagogi, isti program in letnik 20](#_Toc532026418)

[D1. Različni predmeti,različni pedagogi in različni programi 21](#_Toc532026419)

[E0. Medletna primerjava istega predmeta, istega pedagoga, istega programa in letnika 21](#_Toc532026420)

[F0 Primerjava programov v sitem letu 22](#_Toc532026421)

[F1 Medletna primerjava programa 22](#_Toc532026422)

[4. Zaključek 23](#_Toc532026423)

# 1. Uvod

# 2. Statistično metodološki vidiki interpretacij

## 2.1 Predstavitev vprašanj in v študentskih anketah UL

Razprave o lestvicah, ki se uporabljajo pri merjenju stališč, imajo skoraj stoletno tradicijo. Prvi pristopi so bili namreč formalizirani leta 1927 (Thurstenova lestvica) in 1932 (Likertova lestvica). Gre za tako imenovane ordinalne lestvice, kjer je mogoče kategorije odgovorov razvrstiti glede na določen kriterij samega merjenja (npr. se sploh ne strinjam, se ne strinjam, niti-niti, se strinjam, se povsem strinjam). Vse od takrat potekajo obsežne razprave o najrazličnejših metodoloških vprašanjih tovrstnih lestvic. V pogledu oblikovanja vprašanj so najbolj izpostavljeni vidiki:

* število kategorij (2, 3, 4, 5, 6, 7, 8, 9, 10 ...);
* polarnost lestvice (unipolarnost – bipolarnost): biplarna lestvica ima pozitivni in negativni pol, (npr. zgoraj navedeni primer strinjanja), unipolarna pa zgolj stopnjevanje (majhna teža - srednja teža - velika teža),
* narava merjene dimenzije: splošne lestvice (npr. splošno strinjanje s trditvijo »Pedagog ima korekten odnos … se sploh ne strinjam, se ne strinjam, niti-niti, se strinjam, se povsem strinjam«) ali vsebinsko specifične lestvice (npr. »Odnos pedagoga je ….. zelo nekorekten, nekorekten, niti-niti, korekten, zelo korekten«);
* orientacija kategorij odgovorov: naraščajoče ali padajoče;
* format postavitve kategorij odgovorov: vodoravno ali navpično;
* označevanje kategorij: besedilno ali številsko ali oboje ali pa morda grafično (smeški, zvezdice);
* možnost nevsebinskih odgovorov (npr. ne vem) ali njihova izključitev,
* možnost zavrnitve oziroma neodgovora na vprašanje ali prisiljevanje v odgovore (sicer respondent ne more naprej v anketi),
* možnost zavrnitve celotne ankete ali prisiljevanje v sodelovanje (v nasprotnem primeru anketiranec ne more izvesti določene akcije, npr. prijave izpita).

V primeru študentske ankete UL so bile odločitve v odgovarjajočem [Pravilniku UL](file:///C%3A%5CGoogle%20Drive%5C%E2%80%A2%09izra%C4%8Duanavajo%20se%20povpre%C4%8Dja%20za%20programe%20in%20fakultete)[[1]](#footnote-1) naslednje:

* izbrana je bila lestvica s **petimi kategorijami**, ki jo stroka v primeru splošnih lestvic tudi najbolj priporoča.
* uporabi se **bipolarna** lestvica namesto unipolarne;
* izbrana je bila **splošna lestvica** in posledično imajo vsa vprašanja enake kategorije odgovorov dimenzije strinjanja (se sploh ne strinjam, se ne strinjam, niti-niti, se strinjam, se povsem strinjam). Dodati pa velja, da se v stroki po drugi strani kopičijo dokazi, da so splošne lestvice manj kvalitetne od specifičnih.
* Možni odgovori so podani **vodoravno**, v **naraščajočem** redu in **opisno (besedilno), brez številk**. Šele v fazi obdelave se odgovore oštevilči (1, 2, 3, 4, 5); v tem okviru sta torej 1 in 2 negativni vrednosti, vrednost 3 je nevtralna (niti – niti), vrednosti 4 in 5 pa sta pozitivni, s čimer se omogoča srednji odgovor, ki ne preferira nobene od pozitivne oziroma negativne skrajnosti..
* Omogočena je **možnost nevsebinskih kategorij (ne vem, ne želim odgovoriti)** in tudi možnost, da na določeno vprašanje študent enostavno ne poda nobenega odgovora in torej **možnost da ga preskoči** (neodgovor spremenljivke - item nonresponse), kot tudi **možnost, da študent zavrne sodelovanje v celotnem vprašalniku**.

Vprašanja v anketi – vsi vprašalniki so sicer dostopni na podstraneh spletnega mestai [ul.1ka.si](http://ul.1ka.si/r/5/12/Vpra%C5%A1alniki/Vprasalnik_Evalvacija_pred_izpiti/?&p1=1&p2=4)[[2]](#footnote-2) – so bila torej v naslednji obliki, z enakim kategorijami odgovorov:

*V celoti gledano sem s predmetom zadovoljen* …



Nadvse pomembne statistično-metodološke dileme pri obravnavi lestvic za merjenje stališč se pojavljajo tudi v fazi analize. Navedimo najpomembnejše izzive in odgovarjajoče alternative:

* obravnava morebitnih manjkajočih oziroma ne-vsebinskih odgovorov v statistični analizi: ignoriranje ali eksplicitno prikazovanje ali vstavljanje (imputiranje) nadomestnih vrednosti;
* način poročanja o rezultatih: deleži ali povprečja;
* nivo agregiranja ocen in primerjav (npr. pedagog, program, oddelek, fakulteta);
* domicilnost: poročanje o predmetih in izvajalcih določenega programa (ki je definiran z uradnim predmetnikom in kjer lahko predmete poslušajo tudi študenti drugih programov) ali poročanje o predmetih in izvajalcih, ki so jih poslušali študenti določenega programa (ki lahko izbirajo tudi predmete drugih programov in na tej osnovi lahko nastajajo večje razlike);
* prikazovanje rezultatov tudi s kvantilnimi rangi ali ne.

V tem pogledu [pravilnik UL in Funkcionalne specifikacije](http://ul.1ka.si/r/7/4/Gradiva/Pravilnik_o_studentski_anketi_na_Univerzi_v_Ljubljani/)[[3]](#footnote-3) podajajo naslednje usmeritve:

* v fazi analize in izračunavanja rezultatov se manjkajoče in **ne-vsebinske vrednosti ignorirajo**;
* glavni indikator analize in **poročanja so povprečja**;
* privzeto se izračunavajo **povprečja tudi za programe, in fakultete**, mogoče pa so tudi za katedre, oddelke ipd;
* domicilnosti predmetov in izvajalcev se privzeto ne sledi (ampak le na dodatno zahtev), ampak **analize temeljijo na predmetih in izvajalcih, ki so jih poslušali študenti analiziranega programa**;
* prikazujejo se tudi **kvantilni rangi** glede na položaj v programu in fakulteti,

## 2.2 Mejniki pri interpretaciji povprečjij

V stroki se uporabi povprečij za analizo ordinarnih lestvic vse bolj ugovarja, saj v smislu merske lestvice »stališčne razdalje« med kategorijami odgovorov v resnici niso nujne enake, kar je predpogoj za izračunavanja povprečij. Pri ordinarnih lestvicah je namreč mogoče kategorije odgovorov zgolj razvrsti, pri čemer v razvrstitvi (sortiranju) razdalja med njimi ni pomembna in pogosto tudi ni znana. Tipičen primer je razvrstitev športnikov po vrstnem redu, kjer ni več pomembno, ali je med sosednjimi enotama razlika velika ali majhna; lahko je zgolj stotinka sekunde ali pa več sekund. Večnima tudi razlike (razdalje) med sosednjimi enotami (v sortirni vrsti) niso enake. Podobno lahko velja tudi za razliko med kategorijama odgovorov »popolnoma se strinjam« in »se strinjam«, ki v smislu stopnjevanja stališč ni nujno enaka kot je med kategorijama »se strinjam« in »niti – niti«. Enakost razdalj med sosednjama kategorijama odgovorov je dosežena šele v tako imenovani intervalni lestvici, kjer so razdalje med zaporednimi kategorijami odgovorov enake oziroma jih je mogoče primerjati. Povprečja torej lahko pri ordinarnih lestvicah izračunamo le ob (implicitni) predpostavki, da imamo opraviti z intervalno in ne zgolj z ordiniralo lestvico.

Nadaljnji problem pri uporabi povprečij je tudi zanemarjanje porazdelitve. Kadar pa imamo večje asimetrije ali zelo drugačne porazdelitve (npr. U porazdelitev), uporaba zgolj povprečja ni najbolj ustrezna. Skrajni proti primer nastane, ko polovica študentov poda oceno 1, druga pa oceno 5, povprečje pa je enako 3.0.

Kljub zgoraj navedenim ugovorom pa velja dodati, da so povprečja razmeroma robustna. Porazdelitev v obliki črke »U« namreč pri študentskih evalvacijah praktično ne srečamo in v približku so ocene skoraj vedno porazdeljene približno normalno, seveda z znatno asimetrijo v levo, saj je večina študentski ocen običajno 4 in 5, saj je povprečje ocen predmetov in izvajalcev na prvi in drugi stopnji na celotni UL okoli 4.2. Ker ocene nad 5 niso možne, na drugi strani pa so nižje ocene na levi strani seveda mogoče (1, 2, 3), so posledično skoraj vse porazdelitve bolj ali manj asimetrične v levo. Podobno velja tudi za predpostavko o enakosti razdalj med kategorijami intervalne lestvice, ki ni nujno napačna, in se ji skušamo s konstrukcijo lestvice in opredelitve kategorij čimbolj približati.

Brez izračunov povprečjih, ki temeljijo na (ne povsem preverjenih) predpostavkah o (intervalni) lestvici in (normalni) porazdelitvi, je na osnovi ordinalne lestvice mogoče načeloma izračunavati zgolj deleže za določene kategorije odgovorov ter kvantilov in kvantilnih rangov, kot so npr. mediana, kvartili, decili in centili. Metodološko korektno pa je torej rezultate na ordinalnih lestvicah interpretirati zgolj v smislu deležev in kvantilov, kar se tudi vse bolj uveljavlja.

Na drugi strani imajo povprečja - in predpostavka o intervalni lestvici - celo vrsto prednosti. Najpomembnejše je dejstvo, da omogočajo podrobneje primerjave in analize podskupin, saj za ugotavljanje statistično značilne razlike zahtevajo bistveno (večinoma 10 krat!) manjše vzorce kot pa deleži. Nadvse pomembna je tudi možnost izračunavanja variance, ki nato omogoča standardne multivariatne analize (npr. regresija, faktorska analiza ipd.).

Seveda pa so povprečja tesno povezana z analizami na osnovi deležev, kar posebej velja za povezavo z deležem pozitivnih ocen (vrednosti 4 in 5). Navedeno velja tudi za povezanost glavnih mejnikov za interpretacije povprečij.

Deleži pozitivnih odgovorov se zato v kontekstu študentskih anket vse bolj uveljavljajo za interpretacijo ordinalnih spremenljivk. Za razliko od npr. povprečja 3.2, ki nima povsem jasne neposredne interpretacije - saj respondent kot prvo sploh ni podajal številskega odgovora, niti ni jasno, ali vsi študenti res razumejo npr. vrednost 4 (se strinjam) ali razdaljo med 3 (niti-niti) in 4 (se strinjam) na enako način - je namreč delež pozitivnih odgovorov (kategorije »se strinjam« in »povsem se strinjam« oziroma, prevedeno v številke odgovore, kategoriji 4 in 5) interpretativno zelo jasen: gre za delež študentov, ki ima pozitivno stališče. Če npr. rezultati kažejo, da je manj kot 50% študentov podalo pozitivni odgovor, je to vsekakor zaskrbljujoče, saj v takem primeru večina očitno ni bila zadovoljna in je to zelo jasna metodološka ugotovitev, ki ima tudi nesporno vsebinsko interpretacijo. Pričakuje se in večinoma tudi je tako, da je delež pozitivnih odgovorov (4 in 5) vsaj okoli tri četrtine oziroma, da je večji od 80%.

Podobno pomembno in interpolacijsko jasno vlogo ima tudi delež negativnih odgovorov (vrednosti 1 in 2). Gre za delež enot, ki so bile eksplicitno nezadovoljne. V tem okviru se tudi na področju merjenja organizacijske klime med zaposlenimi v organizacijah pogosto uporablja delež 20%. Če je namreč izrecno nezadovoljnih več kot petina enot, potem to že pomeni resen problem. Od pedagoga, ki je za svoje delo plačan (s strani države ali študentov), pa se pričakuje še nekoliko strožji mejnik in v tem okviru se pogosto omenja delež 10% nezadovoljnih študentov kot meja, kjer že velja postati pozoren.

Ker se študentske ocene UL objavljajo predvsem v obliki povprečij, bomo v nadaljevanju podali nekatere mejnike (benchmarke) za **povprečja**, ki jih bomo za boljšo interpretacijo – na osnovi empirične primerjave porazdelitev za realne ocene 2015/2016 – primerjali z odčitanimi deleži pozitivnih oziroma negativnih odgovorov, ki ta povprečja običajno spremljajo:

* $1.0-2.9$**:** v primeru takega povprečja je več negativnih (1, 2) kot pozitivnih (4, 5) ocen; delež pozitivnih ocen (4, 5) je v takem primeru običajno pod 30 %, negativnih ocen pa nad 30 % – gre za kritično nizke ocene, kjer je nujno potrebna obravnava;
* **3.0** $- 3.4$: večina ocen v takem primeru povprečja praviloma še vedno ni pozitivnih, torej prevladujejo ocene 1, 2 in 3. Da večina ni zmogla pozitivne ocene, je seveda zelo slab znak, ki terja pozornost in ukrepanje. Delež pozitivnih ocen (4, 5) je v takem primeru običajno zgolj 30 – 50 %, delež negativnih ocen pa 20 – 30 %;
* **3.5**$-3.9$**:** gre načeloma za pozitivno oceno in glavnina (modus) ocen ima praviloma sicer vrednost 4, vendar je ocen, ki niso pozitivne (1, 2, 3), še vedno več kot pa odličnih ocen (5). S tem povezani delež pozitivnih ocen (4, 5) je običajno 50 – 80 %, delež negativnih ocen pa je običajno 10 – 20 %. Tako povprečje še vedno pomeni, da gre za nedoseganje osnovnega standarda odličnosti, ki velja na številnih univerzah in pomeni, da se takega pedagog – dokler je v tem območju (ali morda še nižje) – ne zaposli, ne vabi na poletne šole, onemogoči napredovanje ipd. V formativnem pogledu ocena vsekakor sporoča, da obstaja jasen prostor za izboljšave. Res pa je, da se pri nekaterih bolj težavnih ocenjevanjih, kot je npr. zadovoljstvo zaposlenih, v praksi večina povprečjih v dobrih organizacijah nahaja v tem intervalu in zgolj v najbolj odličnih preseže 3.9 (v slabih organizacijah pa je pod 3.5). Pri študentskih evalvacijah pa se pričakuje večje povprečje, saj gre za zelo jasno definirano storitev, za katero so študenti močno zainteresirani in za katero je izvajalec velik poznavalec, hkrati pa izvajalce v celoti kontrolira celoten proces. Skoraj vse univerze oziroma fakultete v tem pogledu tudi dosegajo povprečje, ki je nad 3.9 (seveda pa tako povprečje na nivoju univerze ali fakultete še vedno pomeni, da je npr. četrtina posamičnih predmetov po 4.0). V zahtevnejših univerzitetnih okoljih se že o ocenah pod 4.0 opravi določena akcija (npr. pogovor, analiza, včasih celo kak ukrep). Res pa je, da gre pri ocenah pod 4.0 neredko za slučajno odklanjanje, pa tudi za probleme na strani študentov (npr. določena skupina študentov ima šibkejše predznanje);
* **4.0**$-4.4$**:** glavnina ocen ima v takem primeru praviloma sicer še vedno vrednost 4 (modus), vendar je odličnih ocen (5) tokrat več kot ocen, ki niso pozitivne (1, 2, 3). S tem povezani delež pozitivnih ocen (4, 5) je običajno nad 80 %, delež negativnih ocen pa je pod 10 %. Še vedno pa večina ocen (modus) ni bila 5. Povprečna ocena 4.0 ali več je pogosto še vedno za osnovni (oziroma za ciljni) standard kvalitete pedagoškega dela, ki se zahteva na številnih univerzah;
* **4.5**$-4.7$**:** večina enot (več kot 50 %) je podala oceno odlično (modus je 5), delež pozitivnih ocen (4, 5) je v takem primeru običajno nad 90 %, – gre za indic pedagoške odličnosti in tudi pravilnik UL daje v formalnem smislu pri habilitaciji dodatne točke za pedagoško komponento, če ima izvajalec povprečno oceno 4.5 in več. Po najstrožjih kriterijih odličnosti se odličnost tudi sicer začne šele takrat, kadar ima večina ocen odlično vrednost, to je 5;
* **4.8**$-5.0$**:** praktično vse enote (100 %) so podale pozitivno oceno (4 ali 5) in več kot tri četrtine enot je dalo oceno odlično (5) – gre za pedagoško izvedbo, ki v pogledu SET evalvacije skoraj nima napak.

## 2.3 Slučajna napaka in statistična značilnost

Procesi vzorčenja in procesi merjenja v družboslovju generirajo – poleg sistematičnih napak oziroma pristranskosti (angl. bias) – še določeno slučajno variiranje, kar velja tudi za SET ocene. Pri tem pa velja upoštevati naslednje:

* Pri SET ocenah imamo praviloma opravka s celotno populacijo študentov, zato zaradi navedenega razloga sploh ne potrebujemo statističnega sklepanja iz vzorca na populacijo in posledično tudi ne intervalov zaupanja za ocene populacijskih parametrov (npr. povprečja) ali za preverjanja domnev.
* Kljub temu pa je potrebno upoštevati specifično mersko napako, ki nastaja zaradi nivoja zanesljivost (angl. reliability) anketnih odgovorov. Gre za vprašanje (ne)zanesljivosti, zaradi česar bi študent pri ponovljenem merjenju (npr. čez eno uro ali čez en dan) morda odgovoril nekoliko drugače. Tovrstna napaka je pri vprašanjih o dejstvih (npr. leto rojstva) zanemarljiva, pri ocenjevanju stališč pa je lahko razmeroma velika.
* Literatura konsistentno navaja, da variiranje zaradi (nizke) zanesljivosti odgovorov zavzame pri tovrstnih SET ocenah na petstopenjski bipolarni lestvici okoli 10 – 20 % siceršnje elementarne variance povprečja merjene spremenljivke/indikatorja. Z drugimi besedami to pomeni, da je zanesljivost (angl. reliability) od R=0.8 do R=0.9. V najslabšem primeru zavzame vrednost 0.66, kot to lahko razberemo za lestvice strinjanja iz obsežnih analiz izračunavanja SPQ na osnovi ESS. Preostala variabilnost, ki nastaja v takih primerih, in jo lahko pripišemo merski napaki zaradi nezanesljivih odgovorov, znaša (1-R). V primeru popolne zanesljivosti odgovorov, je seveda R=1.0 in preostala varianca zaradi nezanesljivih odgovorov je enaka nič
* Upoštevati velja empirično dejstvo, da je populacijska varianca (Var) za SET ocene na lestvici 1 – 5 večinoma med 0.25 (za povprečja nad 4.0) in 1.0 (za povprečja okoli 3.0). Posledično je odgovarjajoči standardni odklon (SD) okoli 0.5 za višje ocene in 1.0 za nižje ocene.
* Običajni (95%) »Interval zaupanja« ima sicer pri velikih vzorcih iz neskončne populacije siceršnjo obliko ±1.96×SE. Pri tem imamo pri predpostavki enostavnega slučajnega vzorca (SRS) vrednost standardne napake SE=√Var/√n. Če pa se omejimo le na preostalo variabilnost, ki nastaja zaradi (ne)zanesljivosti odgovorov, izhajamo iz izraza ±1.96×SD√(1-R). Če so vzorci manjši, se t=1.96 lahko tudi zmanjša in pri n=60 imamo t=2.00, pri n=30 je t=2.04, pri n=18 imamo t=2.10, pri n=11 je t=2.20, pri n=10, pa 2.2.3.
* Po oceni ekspertov zelo v grobem okoli tretjina izmerjene variance nastaja zaradi merske napake. Seveda pa je obseg odvisen od narave spremenljivke. Pri oceni 1 ali 5 bo merska napaka verjetno manjša kot pri oceni 4. Slednja bi v ponovljenem merjenju lahko bila 3 ali 5, ocena 1 ali 5 pa je tako ekstremna, da bi se tudi v ponovljenem merjenju ohranila bolj verjetno kot pa ocena 4. Pri majhnih n (n<30) je sicer treba namesto 1.96 upoštevati nekoliko večje vrednoti t (t=2.2), vendar to pri zaokroževanju na eno decimalko ne spremeni navedenih vrednosti. Pri oceni npr. povprečja 4.0, kjer je Var=1, R=0.8 in **N=10**, je odgovarjajoča vrednost standardne napake ±2.23×√0.2/√10 =0.3, pri R=0.9 pa ±0.2, pri R=.67 pa ±0.4. Pri N=10 je torej pri običajnih pogojih (**R>0.8)** interval zaupanja večinoma široko za margino okoli ±0.2, npr. 4.0 ± 0.2. V skrajnem primeru **R=0.67** pa se margina v primerjavi z R=0.8 povečajo za okoli četrtino, se pa v primeru R=0.9 v primerjavi z R=0.8 povečajo za več kot za četrtino. Pri večjih skupinah enot se intervali seveda zožijo, npr. pri N==40 pa je ta interval praviloma za dvakrat ožji (za faktor √(N/10)), torej 4.0 ± 0.14, pri N=160 pa še za dvakrat, torej 0.07. Pri N=20 pa je margina v takem primeru enaka 0.2.
* Pri primerjavah dveh predmetov na dveh neodvisnih vzorcih je zato lahko statistično značilna že razlika, ki je za faktor √2=1.4 širša od zgornjih intervalov zaupanja. V primeru (N=10, R=0.9) širine 0.2, je torej statistično značilna razlika 0.3 (npr. 4.0 🡪 4.3)v primeru R=0.8 pa . Pri večjem N se razlika, ki je statistično značilna, v primerjavi s situacijo pri N=10,ustrezno nadalje zoži, in sicer za odgovarjajoči faktor √(N/10). V primeru N=40 je v takem primeru značilna že razlika za 0.015, torej že npr. 4.0 🡪 4.2, kar velja tudi v primeru R=0.8, v skrajnem primeru R=0.67 (kar je malo verjetno) pa bi bila statistično značilna razlika 0.025, torej 4.0 🡪 4.3
* Če obravnavamo predmet oziroma pedagoge na podobnih ali celo istih populacijah, se intervali zaupanja dodatno zožijo. V najbolj skrajnem primeru, ko primerjamo komponente istega predmeta ali primerjamo predmete z istimi študenti, imamo opravka s preverjanjem domnev v odvisnih vzorcih oziroma populacijah, zato se intervali zaupanja zožijo. Empirični podatki za spremenljivke študentskih evalvacij kažejo, da gre v takem primeru za najmanj faktor 2-krat. V takem primeru je statistično značilna že razlika okoli 0.2 in torej pri povprečni oceni (npr. 4.0 🡪 4.2), že pri N=10. in R=0.8. Navedeno velja tudi npr. za primerjave predavateljev na določenem dogodku, kjer so enote, ki ocenjujejo za vsa predavanja, iste. Če pa je število udeležencev večje in/ali je zanesljivost boljša (R>0.8), potem v primeru, kjer so npr. enega predavatelja ocenili 4.1, drugega pa 4.2, gre že pri taki razliki zelo verjetno že za statistično značilno razliko.
* Statistična značilnost seveda še ne pomeni, da so razlike tudi pomembne, še manj to pomeni, da je zanje odgovoren pedagog s svojo pedagoško prizadevnostjo, saj na razlike lahko vplivajo številni dejavniki, ki jih pedagog ne kontrolira.

DODATI in INTEPRATIRRATI

TABELA 1. Širina margine intervala zaupanja v odvisnosti od N (10, 20, 30, 60) in R (0,66, 0,8, 0,9).

## 2.4 Sistematična napaka zaradi neodgovorov

Vse zgornje obravnave veljajo ob predpostavki, da študenti niso izraziteje zavračali sodelovanja in da ne-sodelujoči študenti niso posebej drugačni. Neodgovori so načeloma sicer resna omejitev, saj nekatere študije nakazujejo, da so nerespondenti drugačni („slabši“ študenti), ki dajejo (oziroma bi dajali) nižje SET ocene. Vendar pa v primeru SET ocen na UL teh okoliščin nismo zaznali. Nekaj protiargumentov:

* Dejstvo je, da ni nekega posebnega teoretskega razloga (kot npr. obstaja v primeru, ko osebe z visokimi dohodki težje odgovarjajo v anketi o dohodkih), da bi na tej osnovi nerespondenti imeli drugačen nivo SET ocen.
* Delež zavrnitev (zavrnitev je tudi edini možni način neodgovora, saj študenti nimajo druge izbire: ali izrecno zavrnejo ali pa sodelujejo) med študenti je na članicah UL okoli 30 %:
* Izkušnje na FDV kažejo, da padec sodelovanja iz 65 % na 45 % ni spremenil nivoja povprečnih ocen (npr. povprečje 4.2 je ostalo popolnoma nespremenjeno), kar pomeni, da študenti, ki manj odgovarjajo, ne podajajo bistveno drugačnih ocen..
* Ko je bilo vključeno še vprašanje o pogostosti obiskovanj predavanj, so analize na FDV pokazale, da tudi pogostost obiskovanj predavanj (ki je tesno povezana tudi z zavračanjem sodelovanja), ne vpliva na povprečja ocen študentov. Študenti, ki so obiskovali predavanja pogosto, so podali podobne ocene kot študenti, ki so predavanja obiskovali manj pogosto. Študenti, ki predavanj niso obiskovali, praviloma ne odgovarjajo, saj jim ni treba, kar še posebej velja za prejšnji sistem (od koder so tudi navedene ocene), kjer so študenti prostovoljno izbrali, da sodelujejo v anketi, saj ni bilo nobenega pogojevanja s prijavo.
* Upoštevati velja, da so ocene pristranske za produkt deleža nerespondentov ter razlike v ocenah povprečij med respondenti in nerespondenti. Vendar: tudi v primeru, če bi respondenti dali oceno 4.0, nerespondenti pa 3.0 (kar je popolnoma nerealno velika razlika), imamo pri npr. deležu 30 % nerespondentov napako le 0.3 (4.0 namesto 3.7).

Dodati velja, da sistematična napaka oziroma pristranskost (bias) nastane lahko tudi zaradi merske napake (npr. napačno ali nerodno postavljeno vprašanje), vendar lahko upravičeno predpostavimo, da tovrstne napake ni, saj so vprašanja običajna, ki so preverjena v tovrstnih raziskavah.

## 2.5 Relativna pomembnost razlik

V nadaljevanju se torej ukvarjamo z ocenami, za katere predpostavljamo, da niso pristranske, hkrati pa ocene med seboj primerjamo, pri čemer se osredotočamo izključno na statistično značilne razlike. Razlike, ki niso statistično značilne, namreč niso pomembne in tudi ne morejo postati predmet vsebinskih pogovorov o vzrokih za njihovo prisotnost. Kot smo videli, so zaradi specifičnih statističnih okoliščin pri SET ocenah pogosto značilne že razlike za 0.1, torej že desetinke (npr. 4.1 vs. 4.2).

Ko govorimo o pomembnosti, pogosto uporabljamo relativne razlike, torej spremembo glede na splošen nivo (poprečje). Kdaj je določena razlika relativno pomembna, je sicer odvisno od konteksta. Pri gospodarski rasti je npr. pomembna že desetinka odstotka (npr. 2.5 vs. 2.6), saj pomeni milijone ali milijarde evrov. Podobno je pri stopnji brezposelnosti (npr. 7.8 vs. 7.9), saj v velikih državah (npr. ZDA), desetinka odstotka delovne sile pomeni 100,000 brezposelnih oseb. Nekoliko večje relativne razlike, čeprav še vedno zelo majhne, se pojavljajo pri plačah, saj se v stabilnih družbah socialni partnerji pogosto pogovarjajo o povečanju plač ali pokojnin za odstotek ali nekaj odstotkov. Nekaj odstotkov je tudi običajna razlika pri službenem napredovanju. V splošnem pa je pri družboslovnih fenomenih meja za relativno pomembne razlike nekoliko višja, še posebej ko imamo opravka z ocenami in stališči (in ne kvantitativnimi ocenami kot v primeru plač in BDP). Zelo pogosto kot kriterij relativne pomembnosti uporabimo 5 % ali 10 % relativno razliko.

V primeru SET ocen na specifični lestvici od 1 do 5 – seveda v nadaljevanju predpostavljamo, da so obravnavane razlike statistično značilne – velja naslednje:

* Povečanje za desetinko je v specifičnem okviru lestvice 1 – 5 lahko že blizu relativni pomembnosti, saj 0.1 (npr. 4.0🡪4.1) pomeni okoli 3 % vrednosti povprečne ocene, povečanje 4.0🡪4.2 pomeni 5 %; povečanje 4.0🡪4.4 pa že 10 % vrednosti povprečne ocene;
* Povprečja za SET ocene pa se večinoma gibljejo v intervalu 3.5 – 5.0, saj so ocene pod 3.5 kritično nizke in se zato v okoljih, kjer imajo urejene sisteme spremljanja in obravnavanja kvalitete pedagoškega procesa, pojavljajo redko ali pa se sploh ne pojavljajo. V tem okviru zato sprememba za 0.1 dejansko pomeni kar 7 % omenjenega intervala (3.5-5.0); sprememba v povprečju iz 4.0 na 4.4 pa predstavlja skoraj tretjino (27 %) omenjenega intervala.
* V splošnem torej za SET ocene na lestvici 1 – 5 velja, da je relativno pomembna sprememba lahko že sprememba za desetinko (0.1). Seveda ob tem predpostavimo, da je bila pred tem ugotovljena tudi statistična značilnost. Povsem tretja stvar pa je razlaga, zakaj je taka (značilna in pomembna) razlika nastala, oziroma kaj je njen vzrok, saj razlik nikakor ni mogoče kar avtomatično pripisovati pedagoški prizadevnosti pedagoga. Zaradi ne-eksperimentalne narave SET ocen, ki so opazovalni podatki (angl. observational data), namreč pogosto ne kontroliramo zunanjih dejavnikov, ki lahko vplivajo na SET ocene.

# Vsebinski vidiki interpretacije

## 3.1 Pedagoška prizadevnost

V grobem bi lahko rekli – oziroma vsaj želeli interpretirati SET ocene, ki jih včasih v tej smeri tako tudi poenostavljamo – da je glavni vzrok za razlike v ocenah napor samega pedagoga za kvalitetno pedagoško delo, s katerim nato vpliva tudi na SET ocene.

Seveda pa pogosto na ocene vplivajo tudi drugi zunanji dejavniki in všečnost pedagoga. Ne glede vse ostale vplive pa je pedagoška prizadevnost glavni vzvod, s katerim pedagog vpliva na SET ocene. Izraz »pedagoška prizadevnost« služi v tem okviru kot delovni pojem, ki ga v tem kontekstu uporabljamo za vse vidike pedagoških aktivnosti, povezanih z izvedbo določenega predmeta:

* strukturiranje snovi,
* organizacija predavanj,
* usklajenost delov predmeta, vaj oziroma seminarjev,
* način podajanja snovi,
* obseg in način rabe pedagoških pripomočkov,
* kvaliteta prosojnic,
* kvaliteta izročkov,
* izbor, urejenost in dostopnost literature,
* urejenost spletnega mesta,
* uporaba inovativnih pedagoških pristopov,
* stopnja interaktivnosti,
* optimalna uporaba tehnologije,
* nivo posodabljanja snovi z najnovejšimi raziskovalnimi rezultati,
* povezovanje snovi z aktualnimi družbenimi dogajanji,
* pripravljenosti na predavanja in konciznost podajanja,
* razločnost in jasnost nastopa,
* točnost prihoda in odhoda,
* ustreznost izpitnega procesa in pravičnost ocenjevanja,
* kvaliteta ocenjevanja seminarskih nalog,
* ustreznost obsega študentskega bremena (glede na kreditne točke),
* pedagoški žar in entuziazem za določeno problematiko,
* korektnost odnosa do študentov,
* odprtost za razpravo s študenti izven predavalnice.

S tem so povezane tudi morebitne pomanjkljivosti oziroma nesporazumi, ki nastajajo v izvedbi zaradi površnosti, hitenja, nedomišljenega ravnanja in premajhne pozornosti oziroma preskromnega časa, ki je bil namenjen pripravi oziroma izvedbi predmeta.

Na ocene SET pa lahko poleg pedagoške prizadevnosti vplivajo tudi zunanji dejavniki (npr. število študentov, zanimivost predmeta, specifičnost generacije, specifična infrastruktura, povezana z določeno izvedbo) in všečnost pedagoga (npr. izgled, osebni šarm, komunikativnost). Navedeno v splošnem nima odločilnega vpliva in le redko preseže razlike v obsegu 0.1, 0.2 ali 0.3. Obstajajo sicer določeni tudi primeri, ko zunanji vplivi dosežejo učinek 0.5 (npr. 3.8 🡪 4.3). Kljub temu lahko z dodatnimi napori pedagoške prizadevnosti praktično vsak pedagog doseže vrednosti 4.0.

## 3.2 Vsebinske interpretacije, dejavniki in problem kavzalnosti

Statistično značilne razlike, ki dosežejo tudi določeno relativno pomembnost, imajo neki svoj vsebinski vzrok. Identifikacija konkretnih vzrokov za nizke ali visoke SET ocene je stvar specifične vsebinske razprave in poznavalci (predvsem pa vpleteni pedagogi) jih večinoma lahko izluščijo; posebej, če gre za relativno pomembne razlike, torej za več kot 5 % (npr. razlike med dvema komponentama istega predmeta 4.0- 4.2) ali celo 10 % (3.9-4.3). Nemalokrat pa se soočamo tudi s situacijami, ko je za obstoječe razlike težko poiskati vzroke.

Zavedati se je treba, da se pri SET ocenah soočamo z osnovnim problemom preučevanja vzročnosti (kavzalnosti) v družboslovju, ki nastaja zato, ker ne moremo izvajati eksperimentov oziroma se izvajajo izredno težko in v omejenem obsegu. Posledica je, da zunanjih dejavnikov ne moremo v celoti kontrolirati. Denimo, da je določeno študijsko leto potekala selitev fakultete v nove prostore ali so potekali študentski nemiri – v takem primeru težko ocenimo vpliv navedenih dejavnikov, saj ne moremo izvesti eksperimenta (zavrteti časa nazaj in izvesti predmet na povsem enak način v povsem enakih siceršnjih okoliščinah, vendar zgolj brez navedenih sprememb ter nato analizirati razliko in ovrednotiti, koliko je k oceni prispeval obravnavani vzrok).

Vsekakor pa velja naslednje: več dejavnikov kot lahko v primerjavah kontroliramo, bolj verjetno razlike pojasnimo zgolj s prizadevnostjo pedagoga. Primerjamo lahko namreč SET ocene za isto leto, iste študente, iste predmete, iste predavatelje, iste programe, isto fakulteto, isti letnik ipd., lahko pa določene vidike spreminjamo.

Kadar vseh zunanjih okoliščin ne kontroliramo, lahko zunanji dejavniki in všečnost vplivajo na SET ocene na način, da tega vpliva ni mogoče ločiti od vpliva pedagoške prizadevnosti. Povsem kontroliramo jih pravzaprav le pri primerjavi komponent določenega predmeta pri izvedbi določenega pedagoga. Čim primerjamo med seboj različne predmete, različne pedagoge, različna obdobja, različne študente itd., narašča število dejavnikov, ki jih ne kontroliramo. Praviloma lahko vsi ti zunanji dejavniki vplivajo nekaj desetink (npr. 0.1, 0.2 ali 0.3). Obstajajo pa tudi izjeme okoliščine, kjer vplivajo dejavniki močneje (npr. za 0.5 ali celo več) na primerjave med pedagogi, predmeti, programi, letniki, študijskimi leti. Tako npr. ima lahko študijsko področje oziroma fakulteta sistematičen vpliv, ki dosega 0.5, kar je razlika med ocenami na fakulteti, kjer so študenti s predmeti najbolj zadovoljni (npr. AG ima povprečje 4.5), ter fakultetami, kjer so študenti najmanj zadovoljni (npr. MS, FS, FSD), kjer je povprečje 4.0). Pri tem velja kot zanimivost dodati, da so razlike pri ocenjevanju izvajalcev bistveno manjše, namesto 0.5 so povprečne ocene izvajalcev med navadnimi fakultetami (AG na eni in MF, FSD in FD) razlikujejo za zgolj 0.3.

Do določene mere lahko tudi v takem primeru – s posebnimi postopki statistične analize (npr. analiza variance, regresija, več-nivojska analiza) – zunanje dejavnike izoliramo. Kljub temu se primerjave SET ocen v takih primerih ne sporočajo le v kontekstu razlik v pogledu na pedagoško prizadevnost, kar bi sicer želeli, ampak je pedagoška prizadevnost v določeni meri pomešana z zunanjimi dejavniki. V tem okviru so zato lahko neposredne primerjave SET ocen in poenostavljene interpretacije do določenih predmetov, pedagogov in programov nekorektne. Najlepše se to pokaže v primerih, ko se določen predmet izvaja na povsem enak način za različne skupine študentov, ki pa ga nato ocenijo zelo različno. V tem primeru razlike izvirajo izključno iz specifične percepcije (npr. pričakovanja, zanimanje, predispozicije, predznanje, relevantnost za njihovo področje), ki jo imajo specifične skupine študentov (posebej glede programov) do obravnavanega predmeta oziroma pedagoga. Nazoren je primer statistike, kjer je npr. pri kulturologih zadovoljstvo s pedagogom 3.5 pri novinarjih pa 4.0, čeprav so bili prisotni pri isti izvedbi. Podobno velike so razlike med študenti Anketne metodologije, ki ga univerzitetni študenti ocenjujejo 4.3, visokošolski pa 3.8, čeprav so imeli skupno izvedbo.

##  Pregled interpretacij dejavnikov za tipične primerjave

Nadaljevanju so sistematično predstavljeni zunanji dejavniki – torej dejavniki izven aktivnosti in prizadevnosti pedagoga –, ki lahko vplivajo na razlike med povprečji za ocene. Vključenih je osem potencialnih dejavnikov, ki so analizirani za najpogostejše načine primerjav.

**Tabela 1: Zunanji dejavniki, ki lahko vplivajo – poleg pedagoga – na študentske evalvacijske ocene, glede na objekt primerjave**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Oznaka primerjave | **OBJEKT PRIMERJAVE** | 1. Komponente\* | 2. Predmet | 3. Pedagog | 4. Izvedba | 5- Leto/generacija | 6. Program | 7. Letnik | 8. Fakulteta |
| A | **Komponente** znotraj istega predmeta pri istem pedagogu |  | º | º | º | º | º | º | º |
| B | **Različni predmeti istega pedagoga z istimi študentom**  |  |  | º | º º | º | º | º | º |
| C0 | Istipredmet in **pedagog** in izvedba, različni programi |  |  |  |  |  |  | º º |  |
| C1 | Isti predmetinisti **pedagog**, različna izvedba, različni programi |  |  |  | º º |  |  | º º |  |
| C2 | Različni predmeti, isti **pedagog**, različni programi v istem letu |  |  |  | º º |  |  | º º |  |
| D0 | Različni **predmeti**, različni pedagogi, isti študenti v istem letu |  |  |  | º º |  |  |  |  |
| D1 | Različni **predmeti**, različni pedagog, različni programi v istem letu |  |  |  | º º |  |  | º º |  |
| E0 | Isti **predmet**, pedagog in program – medletna primera  |  |  |  | º º |   |  |  |  |
| E1 | Različni **predmeti**, pedagog in program – medletna primerjava |  |  |  | º º |  |  | º º |  |
| F0 | Različni **programi** v istem letu |  |  |  | º º |  |  | º º |  |
| F1 | Isti **program** – medletna primerjava |  |  |  | º º |   |  |  |  |
| G0 | Različne **fakultete** v istem letu |  |  |  | º º |  |  | º º |  |
| G1 | Ista **fakulteta**  – medletna primerjava |  | º º | º º | º º |   |  | º º |  |
| G2 | Komponente različnih **fakultet**  – medletna primerjava |  |  |  | º º |   |  | º º |  |

\* Pri komponentah (A) gre izjemoma za primerjavo različnih spremenljivk, pri ostalih postavkah (B-G) pa gre vedno za potencialne učinke dejavnikov na določeno (isto) spremenljivko oziroma komponento.

LEGENDA:

º minimalen potencialen vpliv, ki je praviloma zanemarljiv

º º zelo majhen potencialen vpliv, ki je običajno zanemarljiv

 realen potencialen vpliv

V nadaljevanju je sistematično predstavljenih vseh 14 primerov iz Tabele 1, kjer gre najbolj pogoste objekte primerjav.

### A. Primerjave komponent znotraj izvedbe določenega predmeta

Ko primerjamo ocene za komponent (npr. ustreznost literature in sprotno preverjanje znanja) določenega predmeta, je pogosto (N>10, R>0.8) statistično značilna že razlika v povprečju za eno desetinko, torej za 0.1 (4.1 🡪 4.2), pomembna (relativna pomembnost nad 5 %) pa je že razlika za 0.2 (4.1 🡪 4.3), kar skoraj v vseh primerih (N>10) praviloma tudi statistično značilno.

Tiste komponente določenega predmeta oziroma določenega izvajalca pri tem predmetu, ki so ocenjene nižje, jasno kažejo na možnosti izboljšav, kar usmerja pedagoga na določeno pomanjkljivost oziroma nakazuje, kje in kako izboljšati pedagoški proces.

Pri ocenjevanju predmetov v anketi PRED izpitom literatura in spodbujanje k samostojnosti (povprečje 4.1) je najbolj kritična komponenta ocena študenta, najbolje pa je ocenjena obveščenost (4.4). V anketi PO izpitu je najbolj kritičen »*nivo pridobljenih kompetenc*« (ocena je 3.8), pri ocenjevanju pedagogov pa komponenta ocena študenta, v »*kakšni meri* pedagog *spodbuja kritično razmišljanje*« (povprečje na UL je 4.2), medtem ko je povprečje za *pripravljenost* in za *korektnost* bistveno višje (4.5). Ker gre pri navedenih komponentah za razlike na nivoju UL, to govori tudi o vsebinski in inherentni težavnosti določene komponente. Vsekakor pedagog lažje poskrbi zagotovi, da so študenti zadovoljni v smislu obveščenosti, bistveno težje pa jim pričara občutek, da jih spodbuja h kritičnem razmišljanju. Opisane razlike so zato v veliki meri posledica vsebinskih razlik med komponentami. Dodatne razlike med komponentami pa so seveda lahko tudi učinek fakultete in ostalih dejavnikov. Obstaja pa lahko tudi učinek pedagoga, kjer pri določenem pedagogu npr. ima enako slabo obveščanje v primerjavi s percepcijo korektnosti težje posledice kot pri drugem pedagogu. Gre pa pri ostalih učinkih praviloma zgolj za možnost učinkovanja.

Drugih vplivov zunanjih dejavnikov in všečnosti, ki bi vstopali v primerjave, v tem primeru v splošnem ni. Podobno ni niti slučajnih vplivov, saj kontroliramo vse okoliščine, vključno s splošno pedagoško prizadevnostjo.

Morebitne razlike zato govorijo predvsem o specifični prednosti ali pomanjkljivosti specifične komponente pedagoške prizadevnosti pedagoga.

Edini dejavnik, ki morda nekoliko vstopa, je specifičen način, kako študenti razumejo in doživljajo posamezne komponente oziroma vidike. Vendar to še ne pomeni, da komponenta, ki se je izkazala za bolj kritično, pri izboljševanju pedagoške prizadevnosti tudi dejansko zasluži največjo pozornost.

Če je npr. korektnost pedagoga pri določenem predmetu ocenjena 3.9, ostale komponente pa so ocenjene 4.3, potem je pedagoška prizadevnost v pogledu korektnosti obnašanja za takega pedagoga skoraj brez dvoma njegova šibka komponenta. Posledično velja o njej razmisliti, se po potrebi posvetovati in sprejeti določene ukrepe. Obstaja sicer možnost, da je na celotni fakulteti določena komponenta (npr. korektnost odnosa do študenta) zaradi specifičnega dojemanja, ki ga imajo študenti za različne komponente SET ocen, sistematično ocenjena nižje kot ostale komponente kar pri vseh pedagogih. V takem primeru je treba zgornje zaključke o kritičnosti določenega pedagoga omiliti in upoštevati tudi relativno razmerje ocene pedagoga glede na povprečje celotne fakultete oziroma programa, ker gre očitno za sistemsko vprašanje. Dodamo pa lahko, da razlike med komponentami na določeni fakulteti niso nikjer tako velike, zato zgornji primer razlike (3.9 vs. 4.3) nedvomno govori o kritični komponenti pedagoške prizadevnosti, kar nakazuje pravzaprav že razlika 3.9 – 4.1), saj je to statistično značilna razlika že pri 10 enotah, hkrati pa v relativnem smislu predstavlja relativno razliko 5 %. V bistvu je pomembna tudi že razlika za eno desetinko (0.1) torej npr. 3.9 vs. 4.0), saj je tudi slednja statistično značilna, relativna pomembnost je glede na vrednost 4.0 sicer »le« 2.5 %, vendar pa se je ob upoštevanju dejstva, da je celotna skala praktično zgolj 3.5 – 5.0, treba zavedati, da to predstavlja 7 % celotnega razpona.

V primeru ocenjevanja vabljenega predavanja gosta ali predavanja na konferencah, kjer se predavatelji še posebej pripravijo, pa je razpon še bistveno ožji, saj – vsaj pri večjem številu udeležencev – na eni strani ocene skoraj nikoli ne presežejo 4.8, na drugi strani pa le redko zdrsnejo pod 3.9. Razlike v ocenah komponent za 0.1 (npr. 4.2 in 4.3) so zato praviloma pomembne in velja zanje poiskati vzroke.

### B. Primerjave ocen za različne predmete istega pedagoga pri istih študentih v istem letu

Če gre za povsem iste študente, velja glede statistične značilnosti vse enako kot za primerjavo komponent predmeta (A): statistično značilna je praviloma že razlika za 0.1 (npr. 4.1 🡪 4.2). Če študenti niso povsem isti (npr. zaradi izbirnosti), pa so lahko statistično značilne šele nekoliko večje razlike.

Podobno kot pri primerjavi komponent predmeta tudi v tem primeru ne nastajajo nobeni slučajni vplivi ali zunanje okoliščine, ki bi izvirale iz specifičnosti določene skupine oziroma generacije študentov. Študenti so namreč isti in njihove posebnosti (zaradi generacije ali vpisanega programa) so za vse predmete enake.

­­Morebitne razlike zato govorijo predvsem o specifični prednosti ali pomanjkljivostmi določene komponente (npr. urejenost literature) pedagoške prizadevnosti pedagoga.

* Za razliko od primerjave komponent istega predmeta, pa tokrat lahko na primerjave vpliva tudi različna narava predmeta (npr. aktualnost, zanimivost, vsebina) oziroma percepcija teh značilnost med študenti.
* Vplivajo lahko kakšne zelo specifične zunanje okoliščine (npr. termin izvajanja predavanj).

### C0. Isti predmet, ista izvedba, istega pedagoga pri različnih študentih v istem letu

Gre za zelo zanimive razlike, ki pogosto povzročajo čudenje. Pedagog namreč izvaja povsem enake aktivnosti za vse študente, vendar pa ga različne skupine ocenjujejo drugače.

Zaradi različnih populacij se v tem primeru statistična značilnost pojavi šele pri razlikah za 0.2 (npr. 4.1🡪4.3) ali 0.3 (npr. 4.1🡪4.4).

Pri interpretacijah razlik še vedno velja, da všečnost nima vpliva, saj gre za istega pedagoga. Ravno tako na razlike ne vpliva narava in vsebina predmeta (npr. zahtevnost, zanimivost, aktualnost ipd.), saj gre za isti predmet. Praviloma tudi ne obstajajo nobene zunanje okoliščine.

Razlike zato pojasnjujejo predvsem naslednji dejavniki:

* Pedagoška prizadevnost, skupaj z vsemi specifičnimi prednostmi in pomanjkljivostmi posamezne izvedbe; posebej velja izpostaviti, da pedagog določena področja bolje pozna in ima do njih več entuziazma.
* Posebnosti določene skupine študentov; razlike lahko nastajajo zaradi različnih vpisnih pogojev (npr. omejitev ocen), zaradi različne narave področja (npr. bolj ali manj eksaktno področje), strukture študentov (npr. prevladovanje deklet, večje število tujcev ipd.). V tem okviru se lahko pojavi tudi specifična privlačnost določenih vsebin za specifične skupine študentov (npr. študenti nekaterih programov doživljajo statistiko kot nepotreben tujek, drugi kot koristno vsebino).
* Morebitne zelo specifične zunanje okoliščine (npr. termin izvajanja predavanj).

### C1. Isti predmet, različna izvedba, istega pedagoga pri različnih študentih v istem letu

zanimive razlike, ki pogosto povzročajo čudenje. Pedagog namreč izvaja povsem enake aktivnosti za vse študente, vendar pa ga različne skupine ocenjujejo drugače.

Zaradi različnih populacij se v tem primeru statistična značilnost pojavi šele pri razlikah za 0.2 (npr. 4.1🡪4.3) ali 0.3 (npr. 4.1🡪4.4).

Pri interpretacijah razlik še vedno velja, da všečnost nima vpliva, saj gre za istega pedagoga. Ravno tako na razlike ne vpliva narava in vsebina predmeta (npr. zahtevnost, zanimivost, aktualnost ipd.), saj gre za isti predmet. Praviloma tudi ne obstajajo nobene zunanje okoliščine.

Razlike zato pojasnjujejo predvsem naslednji dejavniki:

* Pedagoška prizadevnost, skupaj z vsemi specifičnimi prednostmi in pomanjkljivostmi posamezne izvedbe; posebej velja izpostaviti, da pedagog določena področja bolje pozna in ima do njih več entuziazma.
* Posebnosti določene skupine študentov; razlike lahko nastajajo zaradi različnih vpisnih pogojev (npr. omejitev ocen), zaradi različne narave področja (npr. bolj ali manj eksaktno področje), strukture študentov (npr. prevladovanje deklet, večje število tujcev ipd.). V tem okviru se lahko pojavi tudi specifična privlačnost določenih vsebin za specifične skupine študentov (npr. študenti nekaterih programov doživljajo statistiko kot nepotreben tujek, drugi kot koristno vsebino).
* Morebitne zelo specifične zunanje okoliščine (npr. termin izvajanja predavanj).

### C2. Različni predmeti istega pedagoga pri različnih študentih v istem letu

Zaradi različnih populacij se v tem primeru statistična značilnost pojavi šele pri razlikah za 0.2 (npr. 4.1🡪4.3) ali 0.3 (npr. 4.1🡪4.4).

Pri interpretacijah razlik še vedno velja, da všečnost nima vpliva, saj gre za istega pedagoga. Razlike pojasnjujejo predvsem naslednji dejavniki:

* Pedagoška prizadevnost, skupaj z vsemi specifičnimi prednostmi in pomanjkljivostmi posamezne izvedbe; posebej velja izpostaviti, da pedagog določena področja bolje pozna in ima do njih več entuziazma.
* Posebnosti določene skupine študentov; razlike lahko nastajajo zaradi različnih vpisnih pogojev (npr. omejitev ocen), zaradi različne narave področja (npr. bolj ali manj eksaktno področje), strukture študentov (npr. prevladovanje deklet, večje število tujcev ipd.). V tem okviru se lahko pojavi tudi specifična privlačnost določenih vsebin za specifične skupine študentov (npr. študenti nekaterih programov doživljajo statistiko kot nepotreben tujek, drugi kot koristno vsebino).
* Narava in vsebina predmeta, ki je lahko bolj ali manj zahteven, zanimiv, aktualen ipd.
* Nekatere zelo specifične zunanje okoliščine (npr. termin izvajanja predavanj).

### D0. Različni predmeti in različni pedagogi, isti program in letnik

Zaradi različnih populacij se v tem primeru statistična značilnost pojavi šele pri razlikah za 0.2 (npr. 4.1🡪4.3) ali 0.3 (npr. 4.1🡪4.4).

Razlike pojasnjujejo predvsem naslednji dejavniki:

* Ker gre za različne pedagoge, ima tokrat všečnost pedagoga lahko določen vpliv.
* Pedagoška prizadevnost, skupaj z vsemi specifičnimi prednostmi in pomanjkljivostmi posamezne izvedbe; posebej velja izpostaviti, da pedagog določena področja bolje pozna in ima do njih več entuziazma.
* Narava in vsebina predmeta, ki je lahko bolj ali manj zahteven, zanimiv, aktualen ipd.
* Nekatere zelo specifične zunanje okoliščine izvedbe (npr. termin izvajanja predavanj).

### D1. Različni predmeti,različni pedagogi in različni programi

Zaradi različnih populacij se v tem primeru statistična značilnost pojavi šele pri razlikah za 0.2 (npr. 4.1🡪4.3) ali 0.3 (npr. 4.1🡪4.4).

Razlike pojasnjujejo predvsem naslednji dejavniki:

* Ker gre za različne pedagoge, ima tokrat všečnost pedagoga lahko določen vpliv.
* Pedagoška prizadevnost, skupaj z vsemi specifičnimi prednostmi in pomanjkljivostmi posamezne izvedbe; posebej velja izpostaviti, da pedagog določena področja bolje pozna in ima do njih več entuziazma.
* Posebnosti določene skupine študentov; razlike lahko nastajajo zaradi različnih vpisnih pogojev (npr. omejitev ocen), zaradi različne narave področja (npr. bolj ali manj eksaktno področje), strukture študentov (npr. prevladovanje deklet, večje število tujcev ipd.). V tem okviru se lahko pojavi tudi specifična privlačnost določenih vsebin za specifične skupine študentov (npr. študenti nekaterih programov doživljajo statistiko kot nepotreben tujek, drugi kot koristno vsebino).
* Narava in vsebina predmeta, ki je lahko bolj ali manj zahteven, zanimiv, aktualen ipd.
* Nekatere zelo specifične zunanje okoliščine izvedbe (npr. termin izvajanja predavanj).

### E0. Medletna primerjava istega predmeta, istega pedagoga, istega programa in letnika

Medletne primerjave komponent istega (nespremenjenega) predmeta pri istem izvajalcu za isto ciljno populacijo (program in letnik) so zelo pomembne in so pogosto v ospredju zanimanja.

Kot smo pokazali, so lahko že pri 10 študentih statistično značilne razlike za 0.2 – 0.3 (npr. 4.1 – 4.3), pri predmetih z več študenti pa so razlike še bistveno manjše.

Glavni dejavnik razlik je še vedno pedagoška prizadevnost oziroma specifične pomanjkljivosti izvedbe. Poleg tega seveda lahko vplivajo tudi zunanje okoliščine, ki imajo tri komponente:

* Posebnost generacije (npr. bolj ali manj kritični študenti, omejitev vpisa, prepisi, tujci, število študentov).
* Specifične zunanje okoliščine na fakulteti (npr. novi prostori, morebitni incidenti ali škandali) ali izven fakultete (npr. študentske demonstracije).
* V primeru, da se predmet oziroma njegova izvedba spremeni (npr. urnik, vsebina ali celo nov program), je tudi to lahko potencialni razlog.
* Pedagoška prizadevnost, skupaj z vsemi specifičnimi prednostmi in pomanjkljivostmi posamezne izvedbe; posebej velja izpostaviti, da pedagog lahko v določenem letu nekaj spremebni. V splošnem je dejavnik pedagoga pri tovrstnih primerjavah prevladujoč.

### F0 Primerjava programov v sitem letu

Ko imamo opravka z agregati SET ocen, je slučajno variiranje še bistveno manjše in enako velja tudi za razpon povprečij. Oboje se namreč nadalje zmanjšuje, saj primerjamo povprečja komponent (npr. sedem komponent predmeta, šest komponent pedagoga), pedagoge (vsak ima v povprečju 2.5 predmeta), programe (imajo lahko več deset predmetov) in fakultete (več sto predmetov), zato so statistično značilne že razlike za desetinko (npr. 4.0 🡪 4.1). Posebej ozek je tudi razpon povprečij. Če je npr. na določeni fakulteti skupno povprečje 4.2, so povprečja programov pogosto v intervalu 3.9-4.4. Razlika za eno odstotno točko zato pomeni kar petino (!) celotnega razpona.

Če med seboj primerjamo **različne programe v istem letu**, imajo lahko poleg pedagoške prizadevnosti pedagogov določen vpliv tudi zunanji dejavniki, predvsem profil študentov, narava programa oziroma predmetov ter izvedbene posebnosti.

### F1 Medletna primerjava programa

Če primerjamo **isti program med dvema letoma**, ima prizadevnost pedagogov še posebej veliko vlogo, saj so študenti podobni (v večletnih programih v veliki meri celo isti, npr. triletni program 67 %, dvoletni 50 %), pedagogi pa tudi praktično isti. Vse to zožuje jakost oziroma posledice slučajnih variiranj ter vpliv zunanjih dejavnikov zgolj na neke res zelo posebne primere (npr. novi prostori, študenti na izmenjavi, politične razmere). Že minimalne razlike imajo zato zelo verjetno neke vsebinske vzroke v spremenjeni prizadevnosti pedagogov.

# Zaključek

1. http://ul.1ka.si/r/7/4/Gradiva/Pravilnik\_o\_studentski\_anketi\_na\_Univerzi\_v\_Ljubljani/ [↑](#footnote-ref-1)
2. <http://ul.1ka.si/r/5/12/Vpra%C5%A1alniki/Vprasalnik_Evalvacija_pred_izpiti/?&p1=1&p2=4> [↑](#footnote-ref-2)
3. http://ul.1ka.si/r/7/4/Gradiva/Pravilnik\_o\_studentski\_anketi\_na\_Univerzi\_v\_Ljubljani/ [↑](#footnote-ref-3)