

Statistički podatci

prof. dr. sc. Ksenija Dumičić

Uvodno je neophodno ukazati na važnost statistike kao znanosti te definirati osnovne pojmove vezane za statističke podatke koji se koriste u ekonomskoj i poslovnoj praksi. Ovo poglavlje definira statistički skup, populaciju i uzorak, te opisuje izvore i vrste statističkih podataka, razlikujući kvalitativne i kvantitativne varijable. Definiraju se statistički nizovi, uz ilustraciju mnogobrojnim ekonomskim i poslovnim primjerima.

1.1. Važnost statistike

Tri su razloga izučavanja statistike u sklopu nastavnih programa visokoobrazovnih ustanova u svim zemljama svijeta.

Prvo, numeričke informacije nas okružuju te o njima imamo potrebu saznati nešto više. Primjerice, slušajući televizijski dnevnik korisno je razumjeti što znači bruto domaći proizvod po glavi stanovnika, stopa inflacije, stopa nezaposlenosti, prosječna plaća. Ili, pri čitanju dnevnih novina dobro je znati što je burzovni indeks, te kako koristiti objavljeni tečaj valuta. Nadalje, valja znati vrijede li rezultati neke televizijske ankete (engl. *tele-voting*), ili, smiju li se podatci prikupljeni anketom na uzorku ispitanika poopćavati ili ne.

Drugo, statističke metode su korisne pri donošenju odluka, kako privatnih, tako i poslovnih, pa svakako utječu na naše živote i na poslovanje. Primjerice, pri ugovaranju kasko i životnog osiguranja, korisno je razumjeti što znači rizik, o čemu ovisi izračun premije osiguranja, te kakve financijske posljedice ima kupnja odjednom ili kupnja na rate. Također, kako bismo donijeli odluku o kupnji uputno je razumjeti deklaracije proizvoda zbog prosječnog sadržaja štetnih tvari i sl.

Treće, poznavanje statističkih metoda pomaže pri istraživanju i kvantitativnoj analizi skupova podataka, a nezaobilazno je u znanstvenim istraživanjima manjeg ili većeg broja pojava, koja su često u odnosima zavisnosti.

Predmet proučavanja statistike kao znanstvenoistraživačke discipline su svojstva elemenata statističkih skupova i podskupova. Istraživati se mogu kvalitativna i kvantitativna svojstva tih elemenata. Pritom, elementi izučavanja mogu biti subjekti ili objekti, odnosno to mogu biti osobe, predmeti, prostorne jedinice ili pojave. Svojstva koja se proučavaju variraju od elementa do elementa te se nazivaju **varijablama** ili **obilježjima**, a **modaliteti** ili **vrijednosti** koje poprimaju zovu se **podatci**. Podatci su **statistički** ako ih ima mnogo, ako variraju te ako su prikupljeni na isti unaprijed planirani način.

Načini prikupljanja statističkih podataka svode se na tri istraživačka nacrt (dizajna), a to su: (1) **nacrt ankete**; (2) **nacrt kontroliranih opažanja**; te (3) **nacrt pokusa (eksperimenta)**.

Samo sustavno i planski prikupljeni podatci prema poznatoj istovrsnoj metodologiji mogu biti međusobno **usporedivi**. Tako ne bi valjalo da se analiziraju podatci koji su djelomično prikupljeni metodom telefonske ankete, a djelomično osobnim intervjuom ili metodom pokusa, jer različiti načini prikupljanja rezultiraju različitim kakvoćom podataka.

U najširem smislu **statistika se bavi** prikupljanjem statističkih podataka prema sustavno razvijenom planu prikupljanja, njihovim uređivanjem (grupiranjem), tabličnim i grafičkim prikazivanjem, te analizom i donošenjem zaključaka.

Razlikuju se **dvije skupine statističkih metoda** koje se primjenjuju u različitim područjima stručne i znanstvenoistraživačke djelatnosti, a to su: **deskriptivne** i **inferencijalne statističke metode**.

- **Metode deskriptivne statistike** sastoje se u primjeni postupaka uređivanja i prikazivanja statističkih podataka te u izračunavanju deskriptivnih broječnih pokazatelja. Najčešće se raspolaže s velikim brojem podataka i njihovo uređivanje se svodi na grupiranje, tablično i grafičko prikazivanje. Nakon toga slijedi numerička analiza korištenjem različitih mjera: srednjih vrijednosti, mjera disperzije, mjera asimetrije i zaobljenosti. Konkretni korišteni pokazatelji mogu biti, primjerice, aritmetička sredina, standardna devijacija, koeficijent varijacije, koeficijent asimetrije i drugo. Pritom se svi dobiveni zaključci o karakteristikama analiziranih podataka odnose isključivo na analizirane podatke bez mogućnosti njihova poopćavanja.
- **Metode inferencijalne statistike** su induktivne jer polaze od podskupa ili uzorka podataka, a zaključci se uz primjenu teorije vjerojatnosti mogu poopćiti na cijeli skup. Ove metode polaze od podataka slučajnog uzorka izabranog iz realne populacije ili pak od podataka iz istraživnog statističkog procesa, čije se realizacije mogu smatrati uzorkom. Polazeći tako od uzorka podataka metodama inferencijalne statistike donose se zaključci o cjelini (populaciji) i to procjenjivanjem nepoznatih karakteristika cjeline (populacije) uz određenu razinu pouzdanosti ili testiranjem pretpostavke o tim nepoznatim karakteristikama cjeline (populacije) uz odabranu razinu značajnosti. Zaključci o nepoznatim karakteristikama populacije donešeni na temelju uzorka podataka nemaju apsolutnu sigurnost, stoga se takvi zaključci mogu poopćiti na cijeli skup samo uz određenu **razinu pouzdanosti** i određenu **razinu značajnosti**.

S jedne strane, statistika pripada području prirodnih znanosti, grani vjerojatnost i statistika. S druge strane, ona pripada i području društvenih znanosti kao npr. grani kvantitativna ekonomija, jer izučava društveno-ekonomska kretanja, gospodarstvo i poslovanje poslovnih subjekata, primjenjuje metodu uzoraka u reviziji, kontroli kvalitete proizvoda, ispitivanju zadovoljstva klijenata te ispitivanju javnog mnijenja anketnim istraživanjima. Statistika je znanost koja se razvija pod jakim utjecajem brzog razvoja informacijsko-komunikacijskih tehnologija (engl. *Information and Communication Technology – ICT*) koje omogućavaju jeftino prikupljanje i brzu obradu mase podataka, te brzo prenošenje rezultata statističke analize izravno do krajnjih korisnika ma gdje

oni bili. Razvoj novih znanstvenih disciplina bliskih statistici, primjerice, rudarenja podataka (dubinska analiza podataka, engl. *data mining*), svjedoči o daljnjem multidisciplinarnom razvoju hibridnih disciplina u kojima je statistika nezamjenjiva.

1.2. Statistički skup, populacija i uzorak

Statistički skup je skup jedinica (elemenata) nad kojima se ispituje jedno ili više obilježja, odnosno varijabli. Broj jedinica konačnog statističkog skupa označava se sa N . Broj jedinica N naziva se **opsegom statističkog skupa**. Statistički skup po svom opsegu može biti konačan ili beskonačan. Statistički skup je stvaran ako jedinice toga skupa realno postoje, a hipotetski je ako su jedinice skupa rezultat primjene simulacijskog modela ili eksperimenta.

Skup podataka jednog obilježja (varijable) vezanih za sve jedinice statističkog skupa naziva se **populacija** ili **osnovni skup**. Populacija može biti, kao i statistički skup, konačna ili beskonačna, stvarna ili hipotetična. Veličina populacije jednaka je veličini statističkog skupa. Označava se sa N i naziva se opsegom populacije ili opsegom statističkog skupa.

Podskup statističkog skupa naziva se **uzorkom**, pa se govori o **uzorku jedinica (elemenata)**. Ponekad se bira podskup podataka iz populacije, pa se govori o **uzorku podataka**, odnosno o dijelu statističke populacije. Uzorci su najčešće konačni, a veličina (opseg) uzorka se označava sa n .

U svakom se statističkom istraživanju **statistički skupovi i uzorci moraju precizno i potpuno definirati**, tj. definiraju se **pojmovno, prostorno i vremenski**.

Pojmovnom definicijom utvrđuje se pripadnost svake jedinice skupu s obzirom na pojam. Pojmovi se moraju definirati. Tako nije svejedno analiziraju li se “gotovi proizvodi” ili “poluproizvodi”; “studenti” ili “redovni studenti”; “punoljetna osoba” ili “osoba s pravom glasa”; “kućanstvo” ili “obiteljsko poljoprivredno gospodarstvo”; “turist” ili “domaći turist”; “stan”, “mjesto boravka” ili “mjesto prebivanja” i slično.

Na internetskim stranicama Državnog zavoda za statistiku Republike Hrvatske (u nastavku DZS RH) pod Metodološkim objašnjenjima Popisa stanovništva, kućanstava i stanova u RH 2011. godine (u nastavku Popis 2011., prema: <http://www.dzs.hr>). nalaze se brojne definicije pojmova, kao primjerice:

- “Kućanstvo” je svaka obiteljska ili druga zajednica osoba koje zajedno stanuju i zajednički troše svoje prihode za podmirivanje osnovnih životnih potreba (stanovanje, hrana, i drugo). Kućanstvom se smatra i svaka osoba koja živi sama (jednočlano kućanstvo).
- Pod “znanjem korištenja internetom” podrazumijeva se znanje sljedećih radnji: korištenje dugmetom za navigaciju (*Back, Forward*), nalaženje internetskih stranica putem tražilica (npr. *Google, Yahoo, Altavista* i sl.), dolazak na internetske stranice upisom adrese ili putem veze (linka), korištenje mapom Favorites (za pohranu ili otvaranje često traženih stranica), postava početne stanice (*Homepage*) i korištenje njome, otvaranje stranice u novom prozoru, zaustavljanje i ponovno učitavanje stranice (*Stop, Refresh*).

- “Nezaposleni” su osobe koje zadovoljavaju sljedeća tri kriterija: a) u referentnom tjednu nisu obavljale nikakav posao za novac ili plaćanje u naturi; b) u posljednja četiri tjedna prije anketiranja aktivno su tražile posao; c) ponuđeni posao mogle bi početi obavljati u iduća dva tjedna.

Prostornom definicijom označava se prostor kojemu jedinice statističkog skupa pripadaju. Tako, nije svejedno misli li se na “tradicionalnu regiju Dalmacije” ili na precizno određeno područje “Splitsko-dalmatinske županije”, na “Grad Zagreb” ili “Zagrebačku županiju”, itd.

Vremenskom definicijom određuje se vremenski interval ili vremenska točka za koju su vezane sve jedinice statističkog skupa. Preciznim definiranjem skupa osigurava se obuhvat jedinica koje mu doista pripadaju.

Uz svaki analizirani skup podataka treba pored definicije navesti i **izvor podataka**, odnosno treba barem ukratko opisati metodologiju njihova prikupljanja. Samo potpuno definirani skupovi podataka mogu biti osnova za statističku analizu i zaključivanje.

1.3. Izvori i vrste podataka

U nastavku teksta razlikuju se i opisuju primarni i sekundarni izvori podataka. Statistički podatci se razvrstavaju prema različitim kriterijima, što je neophodno znati za pravilan odabir statističke metode njihove analize. Uvode se pojmovi mjernih skala, statističke varijable te matrice podataka.

1.3.1. Izvori podataka

Izvori statističkih podataka mogu biti primarni i sekundarni. **Primarni izvori** su izvorna istraživanja i ona rezultiraju primarnim podatcima.

Vrste primarnih istraživanja su: (1) *nacrt ankete*; (2) *nacrt promatranja* i (3) *nacrt pokusa (eksperimenta)*. Kod sva tri nacrt radi se o prikupljanju empirijskih podataka, koje tek treba nakon provjere i uređivanja prikladno prikazati u matricama ili tablicama podataka. Nakon njihovog uređivanja slijedi grafička i numerička analiza. Budući da su navedeni nacrti istraživanja ponekad veoma skupi, preporučuje se prethodno provjeriti jesu li potrebni podatci već prikupljeni i objavljeni.

Sekundarni izvori daju sekundarne podatke. Oni su rezultat djelatnosti drugih istraživača, specijaliziranih institucija, odnosno subjekata. Neki od tih podataka su javni podatci, primjerice dani na internetskim stranicama DZS, Hrvatske narodne banke (HNB), ili drugih institucija, a neki se mogu dobiti na zahtjev ili kupiti po komercijalnoj cijeni. Tako, u RH DZS provodi Anketu o dohotku stanovništva (ADS), koja daje obilje sekundarnih podataka, kao i baza podataka Popisa 2011. Nadalje, raspoložive su i baze međunarodnih organizacija, kao npr. baze podataka EUROSTAT-a, organizacija Ujedinjenih nacija (UN), ili baza Svjetske banke (*World Bank Data*) i sl.

1.3.2. Vrste podataka

Klasifikacija podataka na vremenske i presječne

Statistički podatci koji se odnose na istu pojmovnu i prostornu jedinicu, ali se prikazuju kroz vrijeme su **vremenski podatci** koji čine vremenski niz (engl. *time series data*). Vremenski niz je niz kronološki uređenih podataka.

Ako su podatci vremenskog niza vezani za intervale vremena (godine, kvartale, mjeseci, tjedne, dane, minutne intervale i sl.), govori se o **intervalnim vremenskim podatcima**. Ako se pak vremenski podatci odnose na trenutke, odnosno vremenske točke, primjerice datume, nazivaju se **trenutačnim vremenskim podatcima**.

Primjerice, uložni na štednju su vezani za konac razdoblja (trenutak), npr. za 31.12. u godini, pa gledano tijekom niza godina tvore **trenutačni vremenski niz**. Budući da se podatci Popisa stanovništva, kućanstava i stanova uvijek odnose na kritični trenutak Popisa (prema: <http://www.dzs.hr/Hrv/censuses/census2011/firstres/censusmetod.htm>), vezano za konkretnu zemlju, analizirani su za veći broj popisnih istraživanja za dulji niz popisnih godina te tvore trenutačni vremenski niz. Primjer trenutačnih nizova podataka su i podatci o broju zaposlenika u jednom poduzeću tijekom posljednjih 10 godina, kao i niz mjesečnih podataka o vrijednosti zaliha u jednom poduzeću u posljednjih 12 mjeseci.

Kod trenutačnih nizova zbrojevi vrijednosti nemaju smisla. Upravo suprotno, kod **intervalnih vremenskih nizova** podataka zbrojevi vrijednosti imaju smisla, pa se kaže da ti nizovi imaju svojstvo kumulativnosti. Primjerice, intervalni je niz podataka o vrijednosti prodaje trgovačkog poduzeća po kvartalima, kao i niz podataka o broju proizvedenih komada proizvoda. Kod ovih nizova zbroj vrijednosti ima smisla tumačiti. Tako zbroj vrijednosti proizvodnje za dva kvartala govori o polugodišnjoj proizvodnji, što je veoma korisna informacija.

Ako se podatci odnose na istu pojavu u istom vremenu (vremenskom intervalu ili trenutku), ali se odnose na različite prostorne jedinice (općine ili županije, šire regije ili zemlje) ili se odnose na objekte (poduzeća, kućanstva, obitelji, itd.), nazivaju se **presječnim ili prostornim podatcima** (engl. *cross-sectional data*). Presječni podatci se odnose na opažanja vezana za različite jedinice, subjekte ili objekte u fiksnom vremenu, tj. trenutku ili vremenskom intervalu. Jednostavan primjer presječnih podataka odnosi se na $n = 1000$ uzorka kućanstava u RH 28.4.2011. g., tako da se raspolaže sa $n = 1000$ podataka o broju članova kućanstva, $n = 1000$ podataka o njihovu godišnjem prihodu i slično. Drugi primjer obuhvaća podatke za uzorak veličine $n = 250$ malih poduzeća u Zagrebačkoj županiji na dan 31.12.2013. g. s obzirom na broj zaposlenika ($n = 250$ podataka vezanih za jedan trenutak). Treći primjer čine podatci o broju trgovinskih poduzeća po županijama u RH na dan 30.6.2013. g. ($N = 21$ podatak).

Presječni podatci se razlikuju od **longitudinalnih** ili **panel-podataka**, kod kojih postoji više opažanja za svaku jedinicu kroz vrijeme na temelju ponovljenih longitudinalnih istraživanja. Jedinice **panel-uzorka** koji se istražuje mogu biti, primjerice, kućanstva, osobe ili maloprodajni dućani, a njih se ponovljeno ispituje, mjeri ili ih se kontrolirano promatra te se opažanja bilježe. Primjerice, u nekim longitudinalnim anketnim istraživanjima s pomoću standardiziranog anketnog upitnika ispituje se panel-uzorak fiksne veličine od n ispitanika. Takve ankete se mogu provoditi za potrebe primijenjenih društvenih i marketinških te istraživanja za potrebe službene statistike. Tako, DZS godišnje provodi Anketu o dohotku stanovništva (ADS). Istraživanje ADS provodi se na panelu slučajno izabranih privatnih kućanstava (usporediti s Metodološkim objašnjenjima na internetskoj stranici DZS). Ovom anketom se prikupljaju podatci o bruto i neto dohotku

kućanstava i svih članova kućanstava, podatci o obrazovnom statusu osoba, statusu u aktivnosti i zaposlenosti, brizi o vlastitom zdravlju i najmlađim članovima kućanstva, podatci o financijskom i materijalnom statusu kućanstva te podatci o ostalim aspektima životnih uvjeta kućanstava. Istraživanje ADS uvedeno je u statistički sustav RH 2010. godine i to harmonizirano s istraživačkom metodologijom Statističkog ureda Europske unije – Eurostata za istraživanje *EU-SILC* (engl. *Statistics on Income and Living Conditions*), a svrha istraživanja je prikupljanje podataka za izračun pokazatelja siromaštva (engl. *poverty indicators*) za RH. Referentno razdoblje na koje se odnose podatci *EU-SILC* ankete definirano je uzorkom podataka koji se prikupljaju. Tako je referentno razdoblje za podatke o dohotku, koji su najvažniji dio ovog istraživanja, prethodna kalendarska godina. Za ostale podatke referentno razdoblje je različito, primjerice prethodni mjesec od trenutka anketiranja, prethodnih 6 mjeseci, prethodnih 12 mjeseci, itd.

Izvorni i izvedeni podatci

Ako su podatci izraženi u mjernim jedinicama promatranog obilježja, nazivamo ih *izvornima*. Tako su podatci o bruto plaćama zaposlenika u jednom poduzeću izraženi u kunama ili podatci o broju stanovnika po općinama RH izvorni.

Primjenom računskih postupaka na izvornim podatcima nastaju *izvedeni podatci*, primjerice postotci, indeksni brojevi, koeficijenti i dr. Tako su podatci o visini plaća zaposlenika u poduzećima jedne djelatnosti izvorni podatci, a podatci o prosječnoj bruto plaći po zaposlenom ili postotni udjeli muškaraca u ukupnom broju zaposlenika po poduzećima jednoga grada izvedeni podatci. Naime, dobivaju se primjenom računskih postupaka, odnosno dijeljenjem ukupne vrijednosti bruto plaća i broja zaposlenih za svako poduzeće. Nadalje, broj stanovnika po kvadratnom kilometru je statistički omjerni broj ili relativni broj koordinacije, a nastaje dijeljenjem izvornih podataka dviju koordinirajućih veličina.

1.3.3. Mjerne skale i statističke varijable

Statistički podatci su zabilježeni rezultati mjerenja svojstva jedinica statističkih skupova. Cilj statističke analize je ispitati svojstva podataka koja oblikom ili stupnjem variraju od jedinice do jedinice analiziranog statističkog skupa, odnosno uzorka. Ta se svojstva nazivaju *statističkim varijablama* ili *statističkim obilježjima*. Po njima se jedinice statističkog skupa razlikuju, odnosno međusobno sličje. Svako statističko obilježje pojavljuje se u dva ili više oblika ili modaliteta, odnosno kategorije, ili poprma dvije ili više vrijednosti, ovisno o tome je li obilježje kvalitativno ili kvantitativno.

Mjerenje je pridruživanje numeričkih i slovnih oznaka jedinicama statističkog skupa ili uzorka prema određenoj mjernoj skali. Mjerena svojstva variraju, a konkretne varijacije su modaliteti ili vrijednosti statističkog obilježja, odnosno statističke varijable.

Pri mjerenju se koriste *mjerne skale (ljestvice)*. Svaka mjerna skala je skup oblika ili modaliteta statističkog obilježja, odnosno varijable. Oblici se mogu opisati riječima ili su dani kao vrijednosti varijable. Metrička svojstva varijabli, odnosno mogućnosti provođenja aritmetičkih operacija nad modalitetima ili vrijednostima promatrane varijable ovise o vrsti mjerne skale. Glede metričkih svojstava razlikuju se sljedeće četiri vrste mjernih skala:

1) nominalna; **2)** redosljedna (ordinalna ili skala ranga); **3)** intervalna i **4)** omjerna skala.

Kada se modaliteti varijable izražavaju riječima, odnosno slovniim oznakama ili broječanim kodovima, govori se o **kvalitativnim varijablama**, a mjerne skale koje se koriste su nominalna i redoslijedna skala.

1) Nominalna skala je dana u obliku liste pojmova, naziva, odnosno atributa, koji se još nazivaju kategorije ili modaliteti (oblici). Atributi se mogu odnositi na opisno izražen modalitet ispitanog svojstva, kada se govori o atributivnoj skali, ili se odnose na prostorne jedinice, kada se govori o geografskoj skali. Modaliteti mjereni na nominalnoj skali mogu se redati proizvoljno, abecednim redom ili po učestalosti, tj. frekventnosti. Modalitetima se često pridružuju brojčane ili slovne oznake, tj. kodovi (šifre).

Ponekad se nominalna varijabla pojavljuje s vrlo velikim brojem modaliteta pa se primjenjuju **nomenklature**, tj. dogovorno utvrđeni popisi modaliteta nominalne varijable. Nomenklature se koriste kada postoji potreba unificiranog načina navođenja modaliteta nominalne varijable, primjerice, u slučaju potreba službene statistike. Tako, u RH koristi se nomenklatura zanimanja *Nacionalna klasifikacija zanimanja – NKZ-2010*, vidjeti www.dzs.hr, s postavljenim hijerarhijskim sustavom po rodovima zanimanja 0. – 9., pri čemu se svaki rod sastoji od četiriju razina označenih broječanim oznakama prema kojima su razvrstana zanimanja. Nadalje, koriste se još *Nomenklatura industrijskih proizvoda – NIP 2013*, prema www.dzs.hr, nomenklatura gospodarskih djelatnosti nazvana *Nacionalna klasifikacija djelatnosti – NKD-2007*, vidjeti www.dzs.hr, i drugo.

Nad modalitetima nominalne varijable, tj. nad njenim kodovima, nisu dopuštene aritmetičke operacije jer kodovi služe tek kao identifikatori, a modaliteti čine neuređen skup, odnosno mogu stajati u bilo kojem poretku, zbog čega nominalna skala ima najslabija metrička svojstva.

2) Redoslijedna skala (ordinalna ili skala ranga) pridružuje brojeve ili slovne oznake elementima i to prema intenzitetu mjenenog svojstva. Ona služi klasificiranju i uređivanju (redanju) elemenata sukladno stupnju promatranog svojstva. Poredak modaliteta redoslijedne varijable slijedi intenzitet mjenenog svojstva od najjačeg prema najslabijem, ili obrnuto. S modalitetima redoslijedne varijable ne mogu se provoditi aritmetičke operacije, već se mogu samo uspoređivati.

Primjeri redoslijedne varijable su ocjena znanja studenta na ispitu (modaliteti su: nedovoljan, dovoljan, dobar, vrlo dobar, izvrstan); ocjena kvalitete proizvoda (npr., ispravan, neispravan); stupanj stručne spreme (od nekvalificiranog radnika do visoke stručne spreme); stupanj razvijenosti zemlje ili regije (nerazvijena, u razvoju, razvijena, visoko razvijena), i slično.

Statističke varijable koje poprimaju brojčane vrijednosti nazivaju se **kvantitativnim varijablama**, i mjere se na intervalnoj i omjernoj skali.

3) Intervalnom skalom pridružuju se jedinicama statističkog skupa brojevi sukladno intenzitetu mjenenog svojstva. Za tu je skalu karakteristično da ima dogovorno utvrđenu nulu i mjernu jedinicu. Jednake razlike brojeva na intervalnoj skali govore o jednakoj razlici odgovarajućeg mjenenog svojstva. Intervalna skala ima bolja metrička svojstva od prethodno navedenih skala jer omogućuje rangiranje i brojčano utvrđivanje razlika u mjenenom svojstvu, a dopušteno je koristiti i aritmetičke operacije zbrajanja i oduzimanja.

Primjeri intervalnih skala su **temperaturne skale**. Podatci mjereni na intervalnoj skali su: temperatura u Celzijevim stupnjevima i temperatura u Fahrenheitovim stupnjevima. U društvenim istraživanjima stupanj slaganja ili neslaganja osoba s nekim ili s nečim mjeri se s pomoću **Likertove skale**, čiji su modaliteti dani kao: “izrazito se slažem”; “slažem se”; “niti se slažem niti se ne slažem” (nula na Celzijevoj skali); “ne slažem se”; i “izrazito se ne slažem”. Likertova

skala koristi se za mjerenje stavova ispitanika u političkim i nepolitičkim ispitivanjima javnog mnijenja.

PRIMJER 1.1. Celzijeva i Fahrenheitova skala

Mjerenje temperature u Celzijevim stupnjevima ($^{\circ}\text{C}$) ili Fahrenheitovim stupnjevima ($^{\circ}\text{F}$) tipični su primjeri primjene intervalne skale. Kod obiju skala jedinica mjere je udaljenost koja ima konzistentno značenje, a naziva se temperaturni stupanj. Svaki Celzijev stupanj označava $1/100$ udaljenosti između zamrzavanja i ključanja vode, a svaki Fahrenheitov stupanj označava $1/180$ te udaljenosti. Celzijeva skala postavlja nulu na temperaturu smrzavanja vode, dok Fahrenheitova skala ima nulu daleko ispod temperature smrzavanja, pa je $0^{\circ}\text{F} \neq 0^{\circ}\text{C}$. Podatci prikupljeni primjenom ovih temperaturnih skala su kvantitativni te imaju dobra metrička svojstva. Tako, primjerice, ako je u Zagrebu 5°C , a u Splitu 20°C , znači da je u Splitu 15°C više nego u Zagrebu, no ne može se reći da je u Splitu četiri puta toplije nego u Zagrebu.

4) Omjernom skalom se jedinicama statističkog skupa pridružuju vrijednosti (brojevi) sukladno intenzitetu mjenenog svojstva. Varijabla mjerena na omjernoj skali naziva se **numeričkom varijablom**. Za tu skalu je karakteristično da ima definiranu mjernu jedinicu i nulu koja označava nepostojanje ispitivanog svojstva. Tako, primjerice broj djece u obitelji može biti nula, što znači da djece nema, ili broj pritužbi u hotelu može biti nula, što znači da se nitko od gostiju nije žalio. Jednake razlike brojeva na toj skali predočuju jednake razlike mjenenog svojstva, a omjeri vrijednosti ovdje imaju smisla. Podatci dobiveni mjerenjem na omjernoj skali imaju sve karakteristike podataka dobivenih mjerenjem na intervalnoj skali i uz to daju smislene omjere jer se odnose na apsolutnu nulu koja označava potpunu odsutnost mjenene karakteristike. S vrijednostima **numeričke varijable** mjenene na omjernoj skali, moguće je provoditi aritmetičke operacije te stoga omjerna skala ima najbolja metrička svojstva.

Numerička varijabla može biti **diskretna** ili **kontinuirana**. **Diskretne numeričke** varijable primaju konačno ili prebrojivo mnogo vrijednosti. Primjeri diskretne varijable su: broj članova kućanstva, broj zaposlenika u poduzećima, broj postelja u hotelima, broj proizvoda u kutiji, broj nenaplaćenih računa, broj obuće (37, $37\frac{1}{2}$, 38, itd.; ili kod nekih proizvođača 37, $37\frac{1}{3}$, $37\frac{2}{3}$, 38, itd.), broj zemalja sudionica međunarodnog kongresa, i slično. **Kontinuirana numerička** varijabla može poprimiti bilo koju vrijednost iz nekog intervala realnih brojeva. Primjeri ovakve varijable su visina osobe u cm, težina proizvoda u kg, viskoznost tekućine, ali i ekonomske varijable kao visina bruto plaće u kunama, izvoz RH u eurima, cijena proizvoda u kunama, vrijednost prodaje po prodajnim područjima i sl.

PRIMJER 1.2. Kelvinova temperaturna skala

Kod mjerenja temperature u Kelvinovim stupnjevima ($^{\circ}\text{K}$), temperatura od nula stupnjeva (0°K) vezana je za nultu brzinu molekula što uistinu znači da nema temperature, pa je temperatura u tom slučaju mjerena na omjernoj, a ne na intervalnoj skali i smatra se numeričkom varijablom. Vrijedi sljedeće: $0^{\circ}\text{F} \neq 0^{\circ}\text{C} \neq 0^{\circ}\text{K}$, što proizlazi iz različitog položaja dogovornih nula prema Celziju i Fahrenheitu (temperatura je mjerena intervalnom skalom) i potpuno različitog značenja od 0°K , kada temperature nema (prema Kelvinu, temperatura je mjerena omjernom skalom).

Tablica 1.1. Usporedba stupnjeva temperature odabranih točaka prema različitim mjernim skalama

Odabrana točka	Vrsta mjerne skale		
	Omjerna mjerna skala	Intervalna mjerna skala	
	Kelvinovi stupnjevi	Celzijevi stupnjevi	Fahrenheitovi stupnjevi
Apsolutna nula	0,00	− 273,15	− 459,67
Smrzavanje vode	273,15	0,00	32,00
Prosječna temperatura ljudskog tijela	310,00	36,80	98,20
Ključanje vode	373,15	100,00	212,00

Izvor: *Encyclopedia Britannica*

Promjena temperature od jednog Kelvinova stupnja jednaka je promjeni temperature od jednog Celzijeva stupnja. Dvije referentne temperature, točka zaleđivanja vode (0°C) i točka ključanja vode (100°C), odgovaraju redom $273,15^{\circ}\text{K}$, odnosno $373,15^{\circ}\text{K}$, a $0^{\circ}\text{K} = -273,15^{\circ}\text{C}$, tablica 1.1.

Kod Fahrenheitove temperaturne skale jedinica mjere, Fahrenheitov stupanj F određen je dobom razlike između temperature topljenja leda i vrelišta vode, na 180 jednakih intervala. Točka smrzavanja (0°C) odgovara 32°F , a točka vrelišta vode (100°C) je 212°F .

U nastavku, tablica 1.2. pregledno daje usporedbu svojstava pojedinih vrsta statističkih varijabli te navodi niz primjera za svaku, uvažavajući korištenu mjernu skalu uz opis metričkih svojstava (Lind et al., 2008).

Tablica 1.2. Primjeri podataka s obzirom na mjernu skalu

Kvalitativne varijable		Kvantitativne varijable		
Nominalna	Redoslijedna	Intervalna	Omjerna	
Podatci se mogu klasificirati prema modalitetima, koji se redaju abecedno ili prema učestalosti. Računske operacije nisu dozvoljene.	Podatci se mogu rangirati po intenzitetu mjerenog svojstva. Računske operacije nisu dozvoljene. Koriste se operatori $=$, $<$ i $>$.	Jednake razlike brojeva ukazuju na jednaku razliku mjerenih svojstava, a nula je određena dogovorno. Računske operacije zbrajanja i oduzimanja su dozvoljene.	Jednake razlike brojeva ukazuju na jednaku razliku mjerenih svojstava, a nula znači nepostojanje svojstva. Dozvoljene su sve računске operacije.	
Modaliteti (kategorije) su izraženi riječima (kodovima)	Modaliteti (kategorije) su izraženi riječima (kodovima)	Vrijednosti se redaju po veličini	Vrijednosti se redaju po veličini	
Atributivne	Geografske	Temperatura u $^{\circ}\text{C}$ Temperatura u $^{\circ}\text{F}$ Likertova skala (stupanj slaganja ili neslaganja s nekim ili nečim)	Diskretne	Kontinuirane
Spol Boja kose Način plaćanja Zanimanje Djelatnost Pripadnost političkoj stranci	Mjesto rođenja Porijeklo turista Mjesto prebivanja Zemlja porijekla uvezene robe		Ocjena znanja Ocjene kvalitete proizvoda Stupanj zadovoljstva potrošača Stupanj stručne spreme Stupanj ekonomske razvijenosti	Broj bodova na ispitu Broj članova kućanstva Broj učenika u razredu Broj pogrešaka na stranici teksta Broj obuće

PRIMJER 1.3. Podatci i mjerne skale u anketnom istraživanju

Protumačite sadržaj tablice 1.3. u kojoj su prikupljeni podatci iz anketnog istraživanja koje provodi agencija za pružanje usluga u turizmu na uzorku zaposlenika jednog velikog poduzeća koje je klijent spomenute agencije.

- Što su podatci navedeni u stupcima, a što u redovima tablice?
- Komentirajte oznake u prvom retku koje se odnose na pojedine varijable po stupcima s obzirom na njihovu vrstu i metrička svojstva.
- Objasnite što znače brojevi i slovne oznake u poljima tablice.
- Protumačite sve podatke prikupljene za treću jedinicu (trećeg ispitanika).

Tablica 1.3. Podatci za uzorak od $n = 4$ zaposlenika

R. br.	Spol	Navršene godine života	Broj članova kućanstva	Navršene godine radnog staža	Godišnja bruto plaća u EUR	Mjesto godišnjeg odmora: (1= Hrvatska; 2= drugdje u Europi; 3= izvan Europe)	Korištenje usluga <i>touroperatora</i> : 1-da; 2-ne
1	M	43	3	20	25 500,00	3	1
2	Ž	30	2	7	21 000,00	2	2
3	M	31	1	6	19 000,00	1	1
4	Ž	25	2	2	20 500,00	1	1
...							

Rješenje.

- Podatci u redovima odnose se na pojedinog ispitanika u anketi, dok se podatci u stupcima odnose se na pojedine varijable opisane u zaglavlju tablice.
- U prvom stupcu je redni broj jedinice uzorka, tj. ispitanika u anketi. Drugi stupac sadrži odgovore ispitanika vezano za nominalnu atributivnu varijablu “spol” s dva moguća modaliteta (kategorije): muško (kod M) i žensko (kod Ž). Treći stupac sadrži podatke za diskretnu numeričku varijablu “navršene godine života” s cjelobrojnim vrijednostima za svakog ispitanika, koja je mjerena na omjernoj skali. Slijedi četvrti stupac koji sadrži podatke za diskretnu numeričku cjelobrojnu varijablu “broj članova kućanstva”. Podatci za “navršene godine radnog staža” su također numerički diskretni i to cjelobrojni, pa je i ta varijabla mjerena na omjernoj skali. “Godišnja bruto plaća u EUR” je kontinuirana numerička mjerena na omjernoj skali; “mjesto godišnjeg odmora” je nominalna geografska varijabla mjerena na nominalnoj skali s tri ponuđena modaliteta (kategorije). U posljednjem stupcu tablice su nominalni atributivni podatci za varijablu “korištenje usluga *touroperatora*” s ponuđena dva modaliteta (“da”; “ne”), što znači da se radi o dihotomnoj varijabli.
- U poljima tablice nalaze se:
 - vrijednosti numeričkih varijabli (za “navršene godine života” i “broj članova kućanstva”), te
 - slovne i brojčane šifre kvalitativnih varijabli (za “spol”: M = muško; Ž = žensko; za “mjesto godišnjeg odmora”: 1 = Hrvatska; 2 = drugdje u Europi; 3 = izvan Europe) itd.