
OBRADA I ANALIZA PODATAKA

DOBIJENIH ISTRAŽIVANJEM

O značenju izraza „podatak“ 1/3

- Svrha obrade i analize podataka je **crpljenje informacija neophodnih za sticanje odgovora** na pitanje zbog kojeg se istraživanje preduzima.
- Lat. *dare* = dati, *datum*, *data* = ono što je dato, podatak
- „**Podatak**“: upotrebljava se veoma često u raznim kontekstima – **mnoštvo definicija**
- Najčešće se **definiše posredstvom izraza** „**informacija**“, „**činjenica**“ ili „**pokazatelj**“
- Primeri:
 - „Činjenice su događaji ili stanja, a podaci su njihove simbolička reprezentacija.“ (Kneller)
 - „Informacije su podaci koji su obrađeni u oblik koji je smislen za primaoca i koji imaju stvarnu ili opaženu vrednost u sadašnjim ili predstojećim akcijama i odlukama.“ (Dejvis)
 - „Podaci su deskripcije fenomena, tj. diskretne zabeležene činjenice o fenomenima, od kojih dobijamo informacije o svetu.“ (Cičricis)
 - „Podaci su vrednosti pokazatelja koje odgovaraju prihvaćenoj skali.“ (Batigin)
- Neophodnost razmatranja odnosa značenja ovih izraza

O značenju izraza „podatak“ 2/3

- „**Činjenica**“ označava:
 - relativno istinit iskaz o nekom stvarnom stanju stvari
 - samo to stanje stvari
- „**Podatak**“ označava:
 - neki iskaz (ne neophodno istinit)
 - neki znak prirodnog ili veštačkog jezika o stanju stvari (ne neophodno stvarnom)
 - ali ne i samo stanje stvari (podatak ≠ činjenica)
- „**Informacija**“ označava:
 - vrstu podataka (ili obratno: podatak vrstu informacije)
 - neke vrste podataka, a neki podaci neke vrste informacija (logičko preklapanje)
 - upotrebljivi, obrađeni oblik sirove građe podataka
- **Relativnost odnosa**: ono što za jednu osobu može da bude informacija za drugu može biti tek sirovi podatak

O značenju izraza „podatak“ 3/3

- „**Pokazatelj**“ je svojstvo, promenljiva koja može da uzima različite vrednosti iz nekog određenog skupa vrednosti.
- **Podatak je iskaz ili neki znak** (verbalni, brojčani, ikonički ili drugačiji) **nekog jezika** (prirodnog ili veštačkog), **o nekom stanju stvari**, kojim se:
 - tvrdi ili poriče postojanje ili prisutnost nekog objekta, događaja, pojave
 - pripisuje ili odriče neko svojstvo objektu ili neka relacija objekta sa drugim objektima
 - tvrde ili poriču ili upoređuju vrednosti tih svojstava
- **Kriterijumi** za razvrstavanje podataka:
 - **izvor** (npr. primarni i sekundarni – relativna razlika)
 - **priroda** (npr. kvantitativni (razlikovanje kategorija u koje su svrstani objekti) i kvalitativni (razlikovanje kategorija i relacije uređenosti kategorija))
 - **forma** u kojoj su izraženi (npr. verbalni i neverbalni)
- Najznačajnija svojstva podataka: pouzdanost, verodostojnost, tačnost i značajnost
- Podaci:
 - ne nastaju iz ničega, nego su oni plod intencije, oni **se kreiraju**
 - ne mogu da budu sasvim nepristrasni, neutralni i **nezavisni** od vremenskog, kulturnog, teorijskog, metodološkog konteksta

Odabiranje kriterijuma za izbor procedure analize podataka

- Rasprostranjenost upotrebe **statističkih paketa za analizu podataka** (npr., SPSS (*Statistical Packages for the Social Sciences*), SAS (*Statistical Analysis System*), CSS (*Complite Statistical System*), BDMP (*BioMeDical Computer Programs*) i dr.)
- Uz to, neophodno je izabrati odgovarajuće **postupke statističke analize**
- Značajni **kriterijumi** za izbor odgovarajuće procedure analize podataka:
 - **broj promenljivih** koje se istovremeno analiziraju
 - **svrha analize** (deskripcija uzroka ili statističko zaključivanje o osnovnom skupu iz kojeg je uzorak uzet i nivo merenja)
 - **vrsta pretpostavki o rasporedu** osnovnog skupa kojem uzorak pripada
- Prema broju promenljivih uključenih u statistički model, izabiraju se postupci univarijatne, bivarijatne ili multivarijatne analize.
- U grupi postupaka multivarijatne analize razlikuju se dva podskupa:
 - saznavanje međusobne povezanosti promenljivih unutar jednog skupa pri čemu ni jedna promenljiva nije označena kao zavisna od ostalih (faktorska analiza, klasterska analiza i multidimenzionalno skaliranje)
 - saznavanje međusobne povezanosti promenljivih unutar jednog skupa pri čemu je najmanje jedna iz skupa promenljivih određena kao zavisna od drugih
 - postoji samo jedna zavisna promenljiva (višestruka regresija, analiza varijanse i kovarijanse, diskriminaciona analiza i dr.)
 - postoji više od jedne zavisne promenljive (multivarijatna analiza varijanse i kanonička korelacija)

Kriterijumi izbora procedure za analizu podataka

- Teorije statističkog zaključivanja se procenjuju na osnovu sledećih svojstava:
 - vrste informacija
 - vrste teorije verovatnoće i vrste osnovnih načela statističkog zaključivanja
 - prirode rešenja zadataka
- Ovde će biti reči o dve teorije statističkog zaključivanja:
 - klasičnoj (koja filozofsku podlogu ima u usavršenom falsifikacionizmu)
 - bejzijanskoj (koja filozofsku osnovu ima u umerenoj teoriji opravdavanja)

KLASIČNA TEORIJA STATISTIČKOG ZAKLJUČIVANJA

- Informacije su iscrpene iz podataka dobijenih na uzorku osnovnog skupa**
(Prethodne informacije o parametru i informacije o vrednosti posledica mogućih odluka o prihvatanju hipoteze. Informacije se predstavljaju preko statistika ili preko verovatnosti ili funkcije verovatnosti.)
- Usvojeno tumačenje verovatnoće i osnovno načelo statističkog zaključivanja se zasniva isključivo na tumačenju verovatnoće kao relativne učestalosti u dugom nizu.** (Dobijeni podaci predstavljaju jedan od mogućih ishoda, što se naziva – uzoračnim prostorom.)
- Vidovi statističkog zaključivanja**
 - statističko ocenjivanje hipoteza
 - statističko proveravanje hipoteza

Kriterijumi izbora procedure za analizu podataka

KLASIČNA TEORIJA

A. Statističko ocenjivanje hipoteza

Zaključuje se na osnovu podataka dobivenih na uzorku, a da se ne polazi ni od kakve pretpostavke o pravoj vrednosti parametra.

- Ključno pitanje je šta se o nepoznatoj vrednosti parametra može zaključiti na osnovu podataka dobijenih na uzorku, pri čemu je cilj oceniti pravu vrednost parametra.
- **Ocenjivanje** se odnosi na:
 - **brojčane vrednosti parametra** (na osnovu vrednosti izračunate iz podataka dobivenih na uzorku zaključuje o pojedinačnoj vrednosti parametra)
 - **interval** (na osnovu vrednosti izračunate na uzorku zaključuje da prava vrednost parametra pripada nekom intervalu)
- Da bi se na osnovu podataka dobijenih na uzorku ocenila brojčana vrednost parametra, potrebno je **neko pravilo** koje bi za svaki ishod u uzoračnom prostoru kazivalo koju bi vrednost trebalo uzeti kao približnu pravoj vrednosti parametra – **ocenitelj**.
- Postoje različite metode konstruisanja ocenitelja (metod najmanjih kvadrata, metod najveće verovatnosti i dr.)
- Problem izbora o ocenitelja podrazumeva prethodno određivanje merila izbora:
 - **nepristrasnost** (aritmetička sredina uzoračnog rasporeda jednaka je vrednosti parametra)
 - **doslednost** (teženje pravoj vrednosti parametra kada n teži beskonačnom)
 - **efikasnost** (minimalnost varijanse)
 - **dovoljnost** (uzoračni raspored treba da sadrži sve značajne informacije o vrednosti parametra raspoložive u podacima dobijenim na uzorku)

Kriterijumi izbora procedure za analizu podataka

KLASIČNA TEORIJA

B. Statističko proveravanje hipoteza

Polazi se od pretpostavke o pravoj vrednosti parametra.

- Nulta i alternativna hipoteza dele parametarski prostor na dva komplementarna podskupa:
 - **nulta hipoteza** kazuje da prava vrednost parametra pripada nekom određenom podskupu parametarskog prostora
 - **alternativna hipoteza** kazuje da prava vrednost parametra pripada ostatku parametarskog prostora
- Koraci:
 1. Nulta hipoteza (tj. pretpostavljena uzoračna distribucija verovatnoće) izlaže se neposrednoj proveri.
 2. Ako je proveravana hipoteza istinita, različiti ishodi u uzoračnom prostoru imaju različite verovatnoće javljanja.
 3. Te verovatnoće mogu da budu deduktivno zaključene iz distribucije verovatnoće koju navodi proveravana hipoteza.
 4. Kada se dobiju podaci sa uzorka (a oni predstavljaju jedan od mogućih ishoda koji sačinjavaju uzoračni prostor) postavlja se pitanje da li taj ishod iz uzoračnog prostora govori u prilog istinitosti proveravane hipoteze ili protiv nje.
- Raspored verovatnoće o kojem govori proveravana hipoteza deli se na:
 - oblast koju sačinjavaju ishodi čija je pojava vrlo neverovatna (**oblast odbacivanja**)
 - oblast koju sačinjavaju ishodi čija je pojava verovatnija od onih iz oblasti odbacivanja (**oblast prihvatanja**), pri čemu su moguća dva tumačenja:
 - proveravana hipoteza je istinita, ali se ipak pojavio jedan vrlo neverovatni ishod ili
 - manje je verovatno da se (pri uslovu da je proveravana hipoteza istinita) pojavi takav vrlo neverovatni ishod, nego što je verovatno da proveravana hipoteza nije istinita (tumačenje klasične teorije)
- Odluka o odbacivanju ili prihvatanju proveravane (nulte) hipoteze se donosi na osnovu toga da li dobijeni rezultat pripada oblasti odbacivanja ili oblasti prihvatanja.

Kriterijumi izbora procedure za analizu podataka

BEJZIJANSKA TEORIJA

- a) **Informacije koje se koriste u zaključivanju i njihovo izražavanje** (osim informacijama dobijenim na uzorku, koriste se i raspoložive prethodne informacije o vrednosti parametra, koje mogu da se izraze rasporedom verovatnoće na parametarskom prostoru, kako bi se prethodni raspored verovatnoće preinačio u naknadni raspored verovatnoće)
- b) **Teorija verovatnoće na kojoj se zaključivanje zasniva i osnovna načela tog zaključivanja** (Pošto prethodne informacije veoma retko postoje u vidu učestalosti u dugom nizu, prethodni raspored verovatnoće retko može da bude izražen posredstvom verovatnoće kao relativne učestalosti, pa se on izražava tzv. „subjektivnom verovatnoćom“, tj. stepenima uverenosti u nastupanje nekog događaja ili u istinitost nekog iskaza.)
- c) **Vidovi statističkog zaključivanja**
 - statističko ocenjivanje hipoteza
 - statističko proveravanje hipoteza

Kriterijumi izbora procedure za analizu podataka

BEJZIJANSKA TEORIJA

A. Statističko ocenjivanje hipoteza

Zaključuje se na osnovu podataka dobivenih na uzorku, a da se ne polazi ni od kakve pretpostavke o pravoj vrednosti parametra.

- Koriste se i prethodne informacije.
- Ako je odabran odgovarajući verovatnosni model, naknadni raspored verovatnoće sadrži sve što se o pravoj vrednosti parametra može doznati iz podataka dobivenih sa uzorka i raspoloživih prethodnih informacija.
- Pošto funkcija verovatnosti prenosi sve informacije sadržane u podacima dobijenim na uzorku, nema potrebe za razmatranjem uzoračne raspodele, niti za uvođenjem merila nepristrasnosti (značajnih za klasičnu teoriju statističkog ocenjivanja).
- Ocenjivanje bročane vrednosti parametra se obavlja s obzirom na veličinu koja ima najveću naknadnu verovatnoću, a ne očekivanu ili tipičnu vrednost (kao u klasičnoj teoriji).

Kriterijumi izbora procedure za analizu podataka

BEJZIJSKA TEORIJA

B. Statističko proveravanje hipoteza

Polazi se od pretpostavke o pravoj vrednosti parametra.

- Koliko je verovatan neki iskaz da prava, nepoznata vrednost parametra leži u određenoj oblasti parametarskog prostora?
- Ako pođemo od toga da su postavljene dve hipoteze:
 - H_1 : prava vrednost parametra leži u određenoj oblasti parametarskog prostora
 - H_2 : prava vrednost parametra leži u parametarskom prostoru izvan te oblasti
- Iz raspoloživih prethodnih informacija zaključuje se **prethodni raspored verovatnoće**, koji:
 - hipotezi H_1 pridaje verovatnoću $p(H_1)$
 - hipotezi H_2 pridaje verovatnoću $p(H_2)$
- Posle prikupljanja podataka sa uzorka pomoću se na osnovu tih novih informacija preinačuje prethodni raspored verovatnoće u **naknadni raspored verovatnoće** (revidira se prethodna verovatnoća hipoteze).
- Na osnovu dobijenog naknadnog rasporeda verovatnoće – koji o vrednosti parametra kazuje sve što je iz raspoloživih prethodnih informacija i podataka dobijenih (eksperimentom) sa uzorka poznato – mogu da se odrede **naknadne verovatnoće obeju hipoteza**.

Kriterijumi izbora procedure za analizu podataka

BEJZIJANSKA TEORIJA

B. Statističko proveravanje hipoteza

- Pored proveravane (nulte) hipoteze, moguće je uzeti u obzir više alternativnih hipoteza (a ne samo 1)
- Obavlja se **izbor** između:
 - skupine postupaka analize zavisnih od oblika distribucije (**parametrijskih postupaka**)
 - skupine postupaka nezavisnih od oblika distribucije (**neparametrijskih postupaka**)
- Izbor postupaka analize i postupaka statističkog opisivanja i zaključivanja, zavisi od:
 - **nivoa merenja promenljivih koje se analiziraju**
 - **pretpostavki (opštih ili specifičnih) o rasporedu osnovnog skupa** iz kojeg je uzet uzorak
- Mada izraz neparametrijski nije odgovarajući, ovde će tim izrazom biti označeni postupci statističke analize koji zadovoljavaju bar jedan od sledeća dva uslova:
 1. da je promenljiva merena i/ili da se analizovana korišćenjem nivoa merenja nižeg od intervalnog
 2. da se analiza ne zasniva na specifičnim pretpostavkama o rasporedu osnovnog skupa iz kojeg je uzorak uzet (već samo na opštim pretpostavkama kao što su: raspored je neprekidan ili raspored je simetričan).
- Parametrijski postupak (ocenjivanja ili proveravanja hipoteze) se primenjuje:
 - ako je svrha zaključivanje o jednom ili više parametara
 - ako je merenje promenljivih obavljeno na najmanjem intervalnom nivou
 - ako su opravdane specifične pretpostavke o prirodi distribucije
- Parametrijski statistički testovi su: manje robusni i imaju veću statističku snagu.
- Neparametrijski postupak (ocenjivanja ili proveravanja hipoteze) se primenjuje:
 - ako neka od pretpostavki nije održiva, a parametrijski postupak nije robusan u odnosu na nju
- Pošto, za određenu istraživačku situaciju i svrhu istraživanja, postoje odgovarajući alternativni neparametrijski postupci, potrebno je obaviti izbor između njih.

Kriterijumi izbora procedure za analizu podataka

BEJZIJSKA TEORIJA

B. Statističko proveravanje hipoteza

- Pošto je izbor odgovarajućeg postupka statističke analize podataka istraživanja značajan uslov valjanosti statističkog zaključivanja i pošto nisu svi istraživači u području empirijskih nauka znalci statistike, s obrazovnom svrhom je projektovan ekspertni sistem koji u interaktivnom odnosu s korisnikom-istraživačem preporučuje odgovarajuće postupke statističke analize i izgrađena mala baza znanja za prototip takvog ekspertnog sistema
- (Ekspertni sistem je računarski program koji, koristeći se znanjima i procedurama zaključivanja, rešava teške probleme u ograničenom, uzanom domenu i obavlja zadatke za koje se zahteva ekspertiza čoveka.)
- Ekspertni sistem OSPAP (odabiranje statističkih postupaka analize podataka) odabira prikladne postupke univarijatne, bivarijatne i multivarijatne analize podataka dobijenih empirijskim istraživanjem.
- Da bi odabrao odgovarajući postupak, OSPAP istraživaču postavlja niz pitanja:
 - o broju promenljivih koje se istovremeno analiziraju
 - o svrsi statističke analize
 - o nivou merenja promenljivih
 - o prihvatljivosti pretpostavke normalnosti
 - o uzorku (nezavisni ili zavisni)
 - o međuzavisnosti ili zavisnosti promenljivih u slučaju istovremenog analizovanja većeg broja promenljivih
 - o broju zavisnih promenljivih i dr.