

Anuška Ferligoj, Katja Lozar Manfreda, Aleš Žiberna:

OSNOVE STATISTIKE NA PROSOJNICAH

Študijsko gradivo pri predmetu Statistika. Fakulteta za družbene vede, Univerza v Ljubljani

Ljubljana, 2011

1 UVOD V STATISTIKO

1 UVOD V STATISTIKO	1
1.1 KAJ JE STATISTIKA?	2
1.2 OSNOVNI POJMI	3
1.3 VRSTE SPREMENLJIVK	4
1.3.1 Vrste spremenljivk glede na tip izražanja vrednosti	4
1.3.2 Vrste spremenljivk glede na tip merjenja	4
1.4 TIPI STATISTIČNIH ANALIZ	6
1.5 NEKAJ IZ ZGODOVINE STATISTIKE	7
1.6 VAJE	7

1.1 KAJ JE STATISTIKA?

- *Statistika = Sadistika,*
- “*Statistika je laž.*”
- Statistika v vsakodnevnem življenju.
- Statistika ← status = država, politična enota.
- Statistika je veda, ki proučuje množične pojave.
- Statistika se ukvarja z
 - ✓ zbiranjem podatkov (eksperimenti, ankete, štetje ...),
 - ✓ organizacijo, povzemanjem, predstavitvijo, analizo in interpretacijo podatkov,
 - ✓ oblikovanjem veljavnih zaključkov in sprejemanje informiranih odločitev na osnovi statističnih podatkov,
 - ✓ ocenjevanjem sedanosti ali napovedovanjem prihodnosti na osnovi statističnih podatkov.
- 4 pomeni besede **statistika**:
 - 1) Podatki.
 - 2) Značilnosti podatkov, npr. povprečje.
 - 3) Tehnike za zbiranje, analizo in interpretacijo podatkov za podporo odločanju.
 - 4) Znanost, ki razvija in uporablja takšne tehnike.

Zakaj študirati statistiko?

- Ker je obvezen predmet.
- Ker je nepogrešljivo orodje za raziskovanje na vseh znanstvenih področjih.
- Ker je znanje statistike potrebno za razumevanje strokovne literature in pogosto tudi novic.

Obstajajo štiri skupine **uporabnikov statistike**:

- Uporabniki, ki morajo razumeti statistične predstavitve ugotovitev na svojem področju.
- Uporabniki, ki izbirajo, uporabljajo in interpretirajo statistične postopke pri svojem delu.
- Uporabni statistiki (angl. *applied statisticians*).
- Matematični statistiki.

1.2 OSNOVNI POJMI

Enota – posamezni proučevani element.

- redni študent na Univerzi v Ljubljani v študijskem letu 2009/10,
- veliko podjetje (z več kot 250 zaposlenimi), registrirano v Sloveniji in aktivno v letu 2008,
- nogometna tekma v Ligi prvakov v sezoni 2009/10,
- izdelek v proizvodnji nekega podjetja v septembru 2009.

Spremenljivka – lastnost enot; označujemo jih npr. z X , Y , X_I . Vrednost spremenljivke X na i -ti enoti označimo z x_i .

- za študenta: spol, ocena na maturi, izobrazba matere, višina mesečnega dohodka staršev,
- za podjetje: število zaposlenih, povprečni mesečni dohodek zaposlenih, promet v letu 2008,
- za tekmo: kraj igranja, rezultat, število rdečih kartonov, ime domače ekipe, ime gostujoče ekipe,
- za izdelek: vrsta izdelka, prisotnost napake, datum izdelave.

Populacija – množica vseh proučevanih elementov; pomembna je natančna opredelitev populacije (npr. časovno in prostorsko).

- vsi redni študenti na Univerzi v Ljubljani v študijskem letu 2009/10,
- vsa velika podjetja (z več kot 250 zaposlenimi), registrirana v Sloveniji in aktivna v letu 2008,
- vse nogometne tekme v Ligi prvakov v sezoni 2009/10,
- vsi izdelki v proizvodnji nekega podjetja v septembru 2009.

Vzorec – podmnožica populacije, na osnovi katere ponavadi sklepamo o lastnostih cele populacije.

- slučajni vzorec 300 študentov,
- slučajni vzorec 200 velikih podjetij,
- slučajni vzorec 10 nogometnih tekem,
- slučajni vzorec 1000 izdelkov.

1.3 VRSTE SPREMENLJIVK

Matematično: **Spremenljivka** je količina, ki ima različne vrednosti. Nasprotno od **konstante**, ki ima eno samo vrednost.

Statistično: Spremenljivka je značilnost enote.

1.3.1 Vrste spremenljivk glede na tip izražanja vrednosti

1. **Opisne (atributivne) spremenljivke** - vrednosti lahko opišemo le z besedami (npr. poklic, uspeh).
2. **Številске (numerične) spremenljivke** - vrednosti lahko izrazimo s števili (npr. starost, telefonska številka). Med njimi ločimo:
 - **zvezne** (angl. *continuous variable*): teoretično lahko zavzame katerokoli vrednost na nekem intervalu (neskočno število vrednosti). Npr. višina z vrednostmi npr. 172, 172.5, 172.512 (odvisno od natančnosti merjenja).
 - **diskretne** (angl. *discrete variable*): na nekem intervalu lahko zavzame le določeno število vrednosti (končno število vrednosti). Npr. število otrok v družini z vrednostmi 0, 1, 2, 3, ... (ne more biti npr. 2.5, 3.842).

Možne so kombinacije, npr. hišne številke (73A), registerske tablice (LJ D3 274)1.3.2

1.3.2 Vrste spremenljivk glede na tip merjenja

Merjenje: proces pripisovanja vrednosti nekim konceptom.

1. **Nominalne spremenljivke oz. nominalna merska lestvica** (angl. *nominal scale*) – vrednosti lahko le razlikujemo med seboj: dve vrednosti sta enaki ali različni (npr. spol).
2. **Ordinalne spremenljivke oz. ordinalna merska lestvica** (angl. *ordinal scale*) – vrednosti lahko uredimo od najmanjše do največje (npr. stopnja izobrazbe, uspeh, strinjanje z vrednostmi “sploh se ne strinjam”, “ne strinjam se”, “niti niti”, “strinjam se”, “zelo se strinjam”).
3. **Intervalne spremenljivke oz. intervalna merska lestvica** (angl. *interval scale*) – lahko primerjamo razlike med vrednostima dvojic enot; lahko povemo, za koliko vrednosti spremenljivke se neka enota loči od druge (npr. temperatura v °C, koledarsko leto).
4. **Razmernostne spremenljivke oz. razmernostna merska lestvica** (angl. *ratio scale*) – lahko primerjamo razmerja med vrednostima dvojic enot; lahko povemo, kolikokrat večja/manjša je vrednost neke enote od druge enote (npr. starost, čas). Pri razmernostnih spremenljivkah je ničla absolutna ("naravno" določena in ne stvar dogovora kot pri intervalnih).

Različne vrste spremenljivk nosijo različno količino informacij. Rečemo, da imajo različne merske lastnosti.

Vrste spremenljivk od nominalne do razmernostne so urejene od tiste z najslabšimi do tiste z najboljšimi merskimi lastnostmi. Višje merske lestvice (npr. razmernostna) zadoščajo tudi vsem lastnostim nižjih merskih lestvic (npr. nominalne). Ordinalne spremenljivke zadoščajo lastnostim, ki jih imajo nominalne spremenljivke, in podobno razmernostne spremenljivke zadoščajo lastnostim, ki jih imajo intervalne, ordinalne in nominalne spremenljivke.

Intervalne in razmernostne spremenljivke imajo “najboljše” merske lastnosti, ker merjenje temelji na lestvici enakih enot (npr. čas 1 leto, temperatura 1 °C, dohodek 1 EUR, višina 1 cm itd.). Razmernostno spremenljivko spoznamo (za razliko od intervalne), če lahko rečemo:

- da ima neka enota vrednost, ki je k -krat večja od vrednosti neke druge enote (npr. oseba, stara 40 let, ima 2-krat toliko let kot oseba, stara 20 let);
- da je vrednost 0 ena od pomenljivih vrednosti, torej gre za absolutno ničlo (npr. 0 cm je “nič dolžine”, 0 sek je “nič časa”) in ne arbitrarno določena (npr. 0 °C ne pomeni “ni temperature”, leto 0 ne pomeni “ni časa”).

Posamezne statistične metode predpostavljajo določeno mersko lestvico spremenljivk. Npr. povprečja ni smiselno računati za “spol”, pač pa za “dohodek”. Največ učinkovitih statističnih metod je razvitih za intervalne in razmernostne spremenljivke.

1.4 TIPI STATISTIČNIH ANALIZ

Glede na namen statistične analize:

- **opisna statistika** (angl. *descriptive, deductive statistics*) - statistična analiza in opisovanje zbranih podatkov brez težnje, da bi iz teh podatkov posploševali čez njihov obseg;
- **inferenčna statistika** (angl. *inductive statistics, statistical inference*) - statistično sklepanje iz vzorca (dela populacije) na populacijo:
 - ocenjevanje značilnosti populacije (intervali zaupanja),
 - preverjanje domnev (testiranje hipotez).

Glede na število hkrati analiziranih spremenljivk:

- **univariatna** – analiza ene spremenljivke;
- **bivariatna** – analiza dveh spremenljivk;
- **multivariatna** – analiza več spremenljivk.

1.5 NEKAJ IZ ZGODOVINE STATISTIKE

Tri smeri razvoja statistike:

1. Državna statistika
2. Verjetnostna teorija
3. Eksperimentalna statistika

Razvoj statistike v Sloveniji.

Več na statistika.fdvinfo.net, povezava Študijsko gradivo.

1.6 VAJE

- 1) Ferligoj (1994): Naloge iz statistike. Pogl. 2: Osnovni pojmi.
- 2) Za naslednje spremenljivke navedite nekaj možnih vrednosti in povejte, ali so te vrednosti zvezne ali diskretne ter za katero mersko lestvico gre:
 - število poslov, dnevno izpeljanih na ljubljanski borzi,
 - temperatura, ki jo Agencija RS za okolje izmeri za slovenske kraje vsake pol ure,
 - življenjska doba osebnih računalnikov (v mesecih),
 - število letno pripadajočih dni dopusta zaposlenih,
 - dolžine skokov v daljavo na neki atletski tekmi,
 - letni neto dohodek univerzitetnih profesorjev v EUR,
 - velikost bobna pralnega stroja (v kg),
 - število polic knjižnega regala,
 - strinjanje z neko trdivijo na lestvici od 1 do 5, kjer 1 pomeni "sploh se ne strinjam", 5 pa "popolnoma se strinjam".
- 3) Določite, kaj so POPULACIJA, ENOTA, OPREDELJUJOČI POGOJI POPULACIJE, SPREMENLJIVKA in določite VRSTE SPREMENLJIVK (glede na način izražanja vrednosti in glede na mersko lestvico), če želimo proučevati:
 - Starost, spol, strankarska pripadnost, zdravstveno stanje kandidatov za Državni zbor republike Slovenije na zadnjih volitvah.
 - Število knjig po fakultetnih knjižnicah na Univerzi v Ljubljani, 1. oktobra 2004.
 - Uporaba interneta (uporablja/ne uporablja), pogostost uporabe, način dostopa (dial-up, ISDN, ADSL, kabel...) in kraj dostopa med rednimi študenti v Sloveniji v študijskem letu 2003/04.
 - Stališče (za/proti; »sploh se ne strinjam«, »ne strinjam se«, »niti niti«, »strinjam se«, »zelo se strinjam«) do priključitve zvezi Nato v anketi odraslih prebivalcev Slovenije, oktobra 2002.
 - Število na novo diagnosticiranih oseb z aidsom v letu 2001, bruto družbeni proizvod na prebivalca za leto 2001, pripadnost starim ali novim članicam za države EU.

- 4) Sociolog želi preučiti stališča odraslih zaposlenih žensk do subvencioniranja vrtcev. Iz Statističnega registra pridobi naslove slučajnega vzorca 520-ih odraslih zaposlenih žensk, ki jim bo poslal anketni vprašalnik po pošti.
- Kaj je v tej raziskavi populacija?
 - Kaj je v tej raziskavi vzorec in kako velik je vzorec?
 - Zamislite si vsaj štiri anketna vprašanja, na osnovi katerih bi lahko definirali eno nominalno, eno ordinalno, eno intervalno in eno razmernostno spremenljivko.
- 5) Eden od načinov preučevanja stilov leposlovnega pisanja je preučevanje dolžine uporabljenih besed. Študent, ki ga zanima stil pisanja Ivana Cankarja, slučajno izbere eno leposlovno delo tega avtorja ter si zapiše dolžino prvih 250-ih besed v tem delu.
- Kaj je v tej raziskavi enota analize in kaj je spremenljivka?
 - Kakšna je merska lestvica merjene spremenljivke?
 - Kaj je vzorec in kako velik je?
 - Kaj je populacija?
- 6) Primeri izpitni vprašanj
1. Zbrali smo podatke o številu na novo kupljenih knjig leta 2005 v slovenskih knjižnicah in izračunali povprečno število na novo kupljenih knjig. Pri tem je enota opazovanja
- a) ena na novo kupljena knjiga
 - b) ena knjižnica
 - c) število na novo kupljenih knjig
 - d) ena knjiga
2. Zbrali smo podatke o številu na novo kupljenih knjig leta 2005 v slovenskih knjižnicah in izračunali povprečno število na novo kupljenih knjig. Pri tem je spremenljivka
- a) ali je bila knjiga kupljena v letu 2005 ali ne
 - b) ena knjižnica
 - c) število na novo kupljenih knjig
 - d) ena knjiga

REŠITVE:

2)

- število poslov, dnevno izpeljanih na ljubljanski borzi: 223, 268, 355, 177; diskretne, razmernostna
- temperatura, ki jo Agencija RS za okolje izmeri za slovenske kraje vsake pol ure: 17.3, 16.5, 18, 20.2; zvezna, intervalna
- življenjska doba osebnih računalnikov (v mesecih): 18, 18.7, 25, 32; zvezna, razmernostna
- število letno pripadajočih dni dopusta zaposlenih: 22, 25, 28, 30; diskretna, razmernostna
- dolžine skokov v daljavo na neki atletski tekmi: 301, 310, 299, 320; zvezna, razmernostna
- letni neto dohodek univerzitetnih profesorjev v EUR: 19200, 22000, 25000, 21567; diskretna, razmernostna
- velikost bobna pralnega stroja (v kg): 4, 5, 6, 7; diskretna, razmernostna
- število polic knjižnega regala: 1, 2, 10, 7; diskretna, razmernostna
- strinjanje z neko trdivijo na lestvici od 1 do 5, kjer 1 pomeni "sploh se ne strinjam", 5 pa "popolnoma se strinjam": 1, 2, 3, 4, 5; diskretna, ordinalna oz. intervalna

3)

- Starost, spol, strankarska pripadnost, zdravstveno stanje kandidatov za Državni zbor republike Slovenije na zadnjih volitvah.
P: kandidati za DZ RS
E: en kandidat
OPP: volitve v DZ RS, I. 2008
Spremenljivke:
 - Starost: številka, razmernostna
 - Spol: opisna, nominalna
 - Strankarska pripadnost: opisna, nominalna
 - Zdravstveno stanje: merjeno npr. na lestvici zelo slabo, slabo, niti dobro niti slabo, dobro, zelo dobro – opisna ordinalna; merjeno npr. z številom obiskov zdravnika zaradi bolezni v zadnjih 4 letih – številka, razmernostna.
- Število knjig po fakultetnih knjižnicah na Univerzi v Ljubljani, 1. oktobra 2004.
P: fakultetne knjižnice UL
E: ena fakultetna knjižnica UL
OPP: fakultetne knjižnice UL, delujoče in registrirane 1.10.2004
Spremenljivka:
 - Število knjig: številka, razmernostna
- Uporaba interneta (uporablja/ne uporablja), pogostost uporabe, način dostopa (dial-up, ISDN, ADSL, kabel...) in kraj dostopa med rednimi študenti v Sloveniji v študijskem letu 2003/04.
P: redni študenti
E: en redni študent
OPP: redni študenti na fakultetah v Sloveniji, s statusom rednega študenta v štud. letu 2003/04
Spremenljivke:
 - Uporaba interneta (uporablja/ne uporablja): opisna, ordinalna;
 - Pogosto uporabe interneta: merjeno npr. na lestvici »manj kot 1-krat na mesec«, »enkrat mesečno«, »nekajkrat mesečno«, »enkrat tedensko«, »nekajkrat tedensko«, »vsak dan« - opisna, ordinalna.
 - Pogosto uporabe interneta: merjeno npr. z vprašanjem »Kolikor ur ste v prejšnjem tednu uporabljali internet?« - številka razmernostna.
 - Način dostopa (dial-up, ISDN, ADSL, kabel ...): opisna, nominalna
 - Kraj dostopa: opisna, nominalna.
- Stališče (za/proti; »sploh se ne strinjam«, »ne strinjam se«, »niti niti«, »strinjam se«, »zelo se strinjam«) do priključitve zvezi Nato v anketi odraslih prebivalcev Slovenije, oktobra 2002.
P: odrasli prebivalci Slovenije
E: en odrasel prebivalec
OPP: odrasli prebivalci Slovenije, sodelujoči v anketi oktobra 2002
Spremenljivke:
 - Stališče (za/proti): opisna, ordinalna
 - Stališče (»sploh se ne strinjam«, »ne strinjam se«, »niti niti«, »strinjam se«, »zelo se strinjam«) uporablja): opisna, ordinalna oz. intervalna, če odgovorom pripišemo števila npr. od 1 do 5.
- Število na novo diagnosticiranih oseb z aidsom v letu 2001, bruto družbeni proizvod na prebivalca za leto 2001, pripadnost starim ali novim članicam za države EU.
P: države članice EU
E: ena država članica
OPP: članice EU, trenutno
Spremenljivke:
 - Število na novo diagnosticiranih oseb z aidsom v letu 2001: številka, razmernostna
 - BDP/prebivalca za 2001: številka razmernostna
 - Pripadnost starim ali novim članicam: opisna, nominalna.

4)

P: odrasle zaposlene ženske iz Statističnega registra
V: slučajni vzorec žensk, velikost 520 enot

Anketna vprašanja:

Nominalna: »Kje ste zaposleni?«, odgovori »javna uprava«, »zasebni sektor«, »samozaposlena« itd.

Ordinalna: Kakšna je vaša stopnja izobrazbe? Možni odgovori: osnovna šola, 3-letna poklicna šola, 4-letna srednja šola, univerzitetna izobrazba, magisterij, doktorat.

Intervalna: Koliko se strinjate s trditvijo »Država bi morala boljše poskrbeti za subvencioniranje vrtcev«?

Odgovorite na lestvici od 0 do 10, kjer 0 pomeni »Sploh se ne strinjam«, 10 pa »Zelo se strinjam«.

Razmernostna: Kolikšen je bil zadnji mesečni bruto dohodek vašega gospodinjstva?

5)

Enota: beseda

Spremenljivka: dolžina beseda

Merska lestvica: razmernostna

Vzorec: besede iz enega leposlovnega dela, velikosti 250 besed

Populacija: vse besede tega leposlovnega dela

6) 1: b, 2: c