



VJEROJATNOST I STATISTIKA

Upute za izradu projektnog zadatka

Vaš projektni zadatak izradit ćete u programskom jeziku R, a temeljit će se na statističkoj obradi stvarnih i javno dostupnih statističkih skupova na području Republike Hrvatske. Sve podatke možete pronaći u bazi podataka Državnog zavoda za statistiku, a dostupni su putem poveznice web.dzs.hr. Najbitnije informacije pronaći ćete u skripti objavljenoj na Loomenu, ali također dostupnoj na stranicama Sveučilišnog računarskog centra Sveučilišta u Zagrebu.



PC-Axis baze podataka

hrvatski



Odaberi bazu podataka:

Cijene
Industrija
Naselja i stanovništvo Republike Hrvatske
Okoliš
Osobna potrošnja i pokazatelji siromaštva
Poljoprivreda, lov, šumarstvo i ribarstvo
Popis poljoprivrede 2003.
Popis poljoprivrede 2020.
Popis stanovništva 2011.
Pravosuđe
Robna razmjena s inozemstvom
Stanovništvo
Strukturne poslovne statistike
Subnacionalne statistike
Transport i komunikacije
Trgovina na malo
Turizam
Zaposlenost i plaće

Prema temi vašeg zadatka morate odabrati prigodne statičke podatke u arhivi DZS-a (prigodnu bazu podataka) te ih preuzeti na svoje računalo. Na primjer, na donjoj slici smo označili podatke o prosječnim godišnjim neto plaćama u 2000. i 2001. godini, u sektorima “Šumarstvo i sječa drva” i “Ribarstvo”. Primijetite da je potrebno odabrati sve mjesece u godini, ali na način da se označi “I-XII” (u suprotnom ćete dobiti mjesečne, a ne godišnje podatke). Osim izbora sektora i godine, na kraju je potrebno odabrati u kojem ćete formatu izvesti tražene podatke. Najprikladniji format za R programski jezik je .csv datoteka bez zaglavlja te vrijednostima odvojenima znakom zarez. Klikom na “Nastavi” preuzet ćete datoteku sa statističkim podacima.

>> Zaposlenost i plaće >> Tablica 2 Prosječne mjesečne neto plaće

Izaberi tablicu

Izaberi varijablu

Prikaži tablicu

Tablica 2 Prosječne mjesečne neto plaće

Odaberi varijablu

O tablici

Označite vaš odabir i izaberite između tablice na zaslonu i formata datoteke. Savjeti
Za varijable označene * treba odabrati barem jednu vrijednost

DJELATNOSTI * <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Ukupno 103 Označeno 2 Ukupno Poljoprivreda, šumarstvo i ribarstvo Biljna i stočarska proizvodnja, lovstvo i uslužne djelatnosti povezane s njima Šumarstvo i sječa drva Ribarstvo Rudarstvo i vađenje Traži <input type="text"/> <input type="button" value="→"/> <input type="checkbox"/> Početak reda	GODINA * <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Ukupno 14 Označeno 0 2000 2001 2002 2003 2004 2005 Traži <input type="text"/> <input type="button" value="→"/> <input type="checkbox"/> Početak reda
MJESEC * <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Ukupno 13 Označeno 0 I-XII I II III IV V Traži <input type="text"/> <input type="button" value="→"/> <input type="checkbox"/> Početak reda	

Broj označenih ćelija je: 6 (dopušteno je najviše 1.000.000)
Prikaz na zaslonu ograničen je na 1.000 redova i 30 kolona

Razdjeljeno zarezom (csv) bez zaglavlja Nastavi

Nakon što ste preuzeli .csv datoteku, kreirajte radni direktorij u koji ćete je pohraniti i u kojem ćete raditi na vašem kodu. Naredba za određivanje radnog direktorija jest:

```
setwd("putanja do direktorija")
```

Imajte na umu da je notacija putanje do datoteke nešto drugačija u R programskom jeziku nego u Windows OS (skripta, poglavlje 1.3.1.1. Postavljanje radnog direktorija). Nakon toga je potrebno učitati .csv datoteku:

```
df <- read.csv(file="ime_datoteke.csv").
```

Željeni redak odabiremo pomoću naredbe:

```
data <- df$Xime_stupca
```

Imajte na umu da ova naredba neće raditi ako su nazivi stupca jednaki onima u preuzetoj .csv datoteci. Naime, naziv stupca mora biti jedna riječ, ili više riječi povezanih

znakom “_”, ali nikako odvojeni znakom razmaka kao što je slučaj u preuzetoj datoteci (promijeniti naziv stupca iz, npr., “2000 I-XII” u. “2000”).

Za računanje aritmetičke sredine i standardne devijacije koriste se naredbe:

```
mean()  
sd()
```

Vaš zadatak je napisati skriptu koja će u programskom jeziku R za zadane statističke podatke računati osnovne statističke veličine te grafički prikazati podatke koje ste koristili u obradi (stupčasti dijagram). Spremite dijagram u radni direktorij u .png i .jpg format. Veličine koje trebate računati su aritmetička sredina, medijan, donji i gornji kvartil, interkvartil, standardnu devijaciju. Odredite karakteristike distribucije (asimetričnosti i zaobljenost). Postoji li korelacije između dva proizvoljno odabrana razdoblja? Izračunajte Pearsonov i Spearmanov koeficijent korelacije te protumačite rezultate.

Primjer cjelovitog programskog koda koji računa statističke veličine za podatke preuzete u gornjem primjeru:

```
df <- read.csv(file = "DE_RS2_HR.csv")  
data1 <- read_file$X2000  
data2 <- read_file$X2001  
print(paste("Prosječna neto placa u 2000 godini za navedene  
sektore iznosila je", mean(data1), "+-", sd(data1)))  
print(paste("Prosječna neto placa u 2001 godini za navedene  
sektore iznosila je", mean(data2), "+-", sd(data2)))
```

Nakon što ste izračunali sve statističke veličine potrebno je napisati izvještaj u kojem ćete predstaviti vaš projekt. Odredite Vaš statistički skup pojmovno, prostorno i vremenski. Navedite sve veličine koje ste računali, zapišite njihove iznose te ih interpretirajte riječima. Izvješće spremite u .pdf formatu te ga dodajte u radni direktorij. Radni direktorij spremite u komprimiranu mapu naziva “VS_Prezime_Ime.zip” te je unesite u Loomen Sustav ne kasnije od **15. lipnja 2022. U 23:59**. Izrađeni projekt uvjet je za polaganje ishoda učenja i pravo izlaska na usmeni ispit.