



Univerza v Mariboru

Filozofska fakulteta



Univerza v Mariboru

Fakulteta za naravoslovje
in matematiko

KVANTITATIVNE METODE V GEOGRAFIJI - ZBIRKA VAJ Z REŠITVAMI

Danijel Ivajnšič

Maribor, 2020

Kazalo

1 VAJA 1 – PODATKOVNE BAZE

2 VAJA 2 – VZORČENJE

3 VAJA 3 – RELATIVNA ŠTEVILAA (INDEKSI)

4 VAJA 4 – FREKVENČNA PORAZDELITEV ZA ŠTEVILJSKO SPREMENLJIVKO

5 VAJA 5 – KVANTILI

6 VAJA 6 – OKVIR Z ROČAJI

7 VAJA 7 – SREDNJE VREDNOSTI

8 VAJA 8 – MERE VARIABILNOSTI

9 VAJA 9 – NORMALNA PORAZDLITEV (KOLMOGOROV-SMIRNOV PREIZKUS)

10 VAJA 10 – F-PREIZKUS

11 VAJA 11 – T-PREIZKUS

12 VAJA 12 – ANOVA IN KRUSKAL-WALLIS PREIZKUS

13 VAJA 13 – HI-KVADRAT PREIZKUS

14 VAJA 14 – REGRESIJA (LINEARNI TREND)

VAJA 1 – SPLETNE PODATKOVE BAZE

Podatkovna baza, tudi podatkovna zbirka je sklop zbirke dokumentov, medsebojnih sklicevanj na dokumente in sistema za razvrščanje, iskanje in urejanje podatkov v bazi. V geografiji pogosto uporabljamo obstoječe spletne podatkovne baze kot so:

- STAT.SI → spletna baza podatkov Statističnega urada RS <https://pxweb.stat.si/SiStat/sl>
- STAGE → spletna baza prostorskih podatkov Statističnega urada RS <https://gis.stat.si/>
- EUROSTAT → spletna baza podatkov Evropske komisije <https://ec.europa.eu/eurostat/web/main/data/database>
- WORLDBANK → globalna spletna baza podatkov pod okriljem združenja petih inštitucij IBRD, IDA, IFC, MIGA in ICSID <https://www.worldbank.org/>

Naloga:

1. Uporabi aplikacijo STAGE in pripravi tabelo podatkov za slovenske statistične regije v letu 2017 z naslednjimi kazalci:
 - število prebivalcev skupaj
 - indeks staranja skupaj
 - gostota prebivalstva
 - število po izobrazbi (višja in visokošolska izobrazba)
2. Podatke smiselno uredi v tabelo in jih grafično prikaži

VAJA 2 – VZORČENJE

Na splošno podatke pridobivamo na dva načina: z popolnim (popis, tekoča registracija) ali delnim opazovanjem (vzorčenje). V praksi je bolj pogosto vzorčenje, ki se glede na način izbire enot za vzorec deli v neslučajno in slučajno (temelji na določenem verjetnostnem zakonu). Zanima nas predvsem enostavno slučajno vzorčenje brez omejitev v katerega sodita enostavno slučajno vzorčenje in sistematično vzorčenje (pogojno). Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. Na podlagi podatkov o prometnih nesrečah v Sloveniji za leto 2013 izvedi enostavno naključno vzorčenje ($N=30$), metodo 3x ponovi in primerjaj dobljeni vzorec
2. Nadaljuj s sistematičnim vzorčenjem, določi velikost vzorca, velikost populacije, faktor k in izberi vzorec, metodo enkrat ponovi in primerjaj dobljeni vzorec

VAJA 3 – RELATIVNA ŠTEVILA (INDEKSI)

Z indeksi ali indeksnimi števili izražamo za koliko je primerjani podatek v stotinkah večji ali manjši od baze. Med seboj primerjamo istovrstne podatke lahko iz vseh statističnih vrst in relativnih števil. Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. Izračunaj absolutno razliko kazalca emisije toplogrednih plinov (v 1000 t CO₂ ekvivalenta) po izbranih državah v časovni vrsti (1999-2010) pri čemer je baza oziroma primerjalno leto 1999.
2. Izračunaj absolutno razliko kazalca emisije toplogrednih plinov (v 1000 t CO₂ ekvivalenta) po izbranih državah v časovni vrsti (1999-2010) pri čemer je baza oziroma primerjalno leto 1999.
3. Izračunaj indeks kazalca emisije toplogrednih plinov (v 1000 t CO₂ ekvivalenta) po izbranih državah v časovni vrsti (1999-2010) pri čemer je baza oziroma primerjalno leto 1999.
4. Izračunaj verižni indeks kazalca emisije toplogrednih plinov (v 1000 t CO₂ ekvivalenta) po izbranih državah v časovni vrsti (1999-2010)
5. Izračunaj stopnjo rasti kazalca emisije toplogrednih plinov (v 1000 t CO₂ ekvivalenta) po izbranih državah v časovni vrsti (1999-2010)
6. Nariši kombinirani stolpčno-linijski grafikon za 3 izbrane države pri čemer je bo vrednost indeksa prikazana v stolpcih, vrednost verižnega indeksa pa z linijami.

VAJA 4 – FREKVENČNA PORAZDELITEV ZA ŠTEVILJSKO SPREMENLJIVKO

Kadar ima preučevana statistična množica veliko število enot (npr. nad 100), teh enot ne moremo urediti v urejeno statistično vrsto. Ponazoritev statistične množice s statistično vrsto bi bila nepregledna in komplicirana. Kadar se srečamo s tem problemom, uporabimo frekvenčno porazdelitev. Statistične vrste, ki prikazujejo sestavo populacije po skupinah vrednosti številske spremenljivke, to je po razredih, imenujemo frekvenčne distribucije ali frekvenčne porazdelitve (pogostne porazdelitve). Frekvenčno porazdelitev dobimo z razvrstitvijo enot glede na vrednost spremenljivke po razredih. S preštevanjem enot po razredih dobimo frekvenco razreda, ki predstavlja število enot proučevane populacije, katerih vrednost se nahaja znotraj mej razreda. Paziti moramo, da bodo mejne vrednosti vključene natančno v en razred in, da je zagotovljena preglednost podatkov. Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. Oblikovati je potrebno frekvenčno porazdelitev za spremenljivki (a) površina izbranih občin v Sloveniji in (b) povprečna višina padavin (v mm) za izbrane meteorološke postaje. Izberite ustrezno zaporedje za določanje širine razredov in oblikujte popolno frekvenčno tabelo, ki vsebuje naslednje kazalce:
 - število razredov (k)
 - širina razredov (ik)
 - zgornja meja razreda ($Y_{k,z}$)
 - spodnja meja razreda ($Y_{k,s}$)
 - interval razreda (ik)
 - vrednost razreda (frekvenca = fk)
 - sredina razreda (Y_k)
 - strukturni delež ($f'k$)
 - strukturni odstotek ($fk\%$)
 - gostote frekvenc (gi) (v primeru neenake širine razreda)
 - kumulativna frekvenca razreda (Fi)
 - relativna kumulativna frekvenca razreda ($fi\%$)
2. Narišite histogram (frekvenca po razredih).

VAJA 5 – KVANTILI

Ranžirna vrsta = enote spremenljivke urejene v zaporedje od najmanjše do največje vrednosti

Rang (R) = je zaporedno mesto vrednosti v ranžirni vrsti

Relativni rang (P) = relativno, običajno v deležu, izražen R

Kvantil = je vrednost, ki razdeli ranžirno vrsto v dva dela

Mediana (Me) = razdeli ranžirno vrsto na dva enaka dela

Kvartili (Q1, Q2, Q3, Q4) = razdelijo ranžirno vrsto na štiri enake dele

Kvartilni razmik (Q) = je razdalja med tretjim in prvim kvartilom ($Q = Q3 - Q1$)

Decil ($D_{1...10}$) = razdeli ranžirno vrsto v deset enakih delov

Centil ($C_{1...100}$) = razdeli ranžirno vrsto v sto enakih delov

Rang grafikon = na abscisi prikažemo vrednosti spremenljivke, na primarni ordinati vrednosti ranga (R) in na sekundarni ordinati vrednosti relativnega ali kvantilnega ranga (P).

Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. Za spremljavo skupno število prebivalcev po slovenskih statističnih regijah zračunaj:
 - ranžirno vrsto
 - rang (R)
 - relativni ali kvantilni rang (P)
 - mediano (Me)
 - Q1, Q2 in Q3
2. Zanima nas koliko % slovenskih statističnih regij ima manj kot 100000 prebivalcev in koliko manj kot 200000 prebivalce
3. Nariši rang grafikon

VAJA 6 – OKVIR Z ROČAJI

Pri okvirju z ročaji (*ang. Box and Whiskers plot*) gre za grafični prikaz vrednosti pogojnega maksimuma, pogojnega minimuma, vrednosti kvartilov Q1, Q2 in Q3 ter potencialnih osamelcev. Okvir z ročaji prikazuje pogojni maksimum in minimum le v primeru kadar se v vzorcu pojavijo vrednosti, ki ustrezajo pogojem bodisi zgornjega ali spodnje osamelca. Osamelci so definirani na podlagi vrednosti kvartilnega razmika. Opozoriti velja, da obstajajo različni načini računanja mejne vrednosti za osamelce ali ekstremne osamelce.

Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. Za vrednost spremljave, ki se je beležila na štirih lokalitetah (skupine alfa, beta, gama, delta) narišite okvir z ročaji. Uporabite bližnjico za risanje tega grafikona, ki jo ponuja Excel 2013 in višje verzije
2. Za vrednost spremljave, ki se je beležila na štirih lokalitetah (skupine alfa, beta, gama, delta) narišite okvir z ročaji. Tokrat ga narišite s pomočjo naloženih stolpčnih diagramov in orodja za vstavljanje intervalov napak.
3. Oba postopka risanja okvirjev z ročaji ponovite še na primeru podatkov o letni količini vode, dobavljene iz javnega vodovoda v Slovenija (1000 m³), pri čemer primerjate Osrednjeslovensko in Podravsko statistično regijo.

VAJA 7 – SREDNJE VREDNOSTI

Pogosto želimo vrednosti spremenljivke predstaviti z njenimi predstavniki, da bi dobili bolj nazorno predstavo o spremenljivki. Najpogosteje uporabljeni predstavniki spremenljivke so njene srednje vrednosti, ki so mere sredine spremenljivke. Ker lahko sredino spremenljivke vrednotimo na različne načine, je srednjih vrednosti več: modus, mediana in povprečje. Katera srednja vrednost je za določeno spremenljivko primerna, je odvisno od njene merske lestvice. Mere sredine posledično lahko računamo tudi iz podatkov, ki so razvrščeni v frekvenčno porazdelitev.

Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. Spremenljivka povprečna višina padavin (v mm) na izbranih meteoroloških postajah v Sloveniji je že razvrščena v frekvenčno porazdelitev z sedmimi razredi enake širine. Izračunaj naslednje kazalce:
 - Modus (Mo)
 - Mediana (Me)
 - Q1 in Q3
 - Aritmetično sredino
2. Imamo podatke o telesni masi (kg) in višini (cm) za 116 študentov geografije na UM. Za obe spremenljivki izračunaj:
 - Modus (Mo)
 - Mediana (Me)
 - Q1 in Q3
 - Aritmetično sredino
3. Iz podatkov o telesni masi (kg) in višini (cm) za 116 študentov geografije na UM izračunaj indeks telesne mase (ITM). Za dobljene rezultate ITM nato izračunaj:
 - Modus (Mo)
 - Mediana (Me)
 - Harmonično sredino
4. Imamo podatke o časovni vrsti števila štipendistov (v 1000) med letoma 1990 in 1996. Zanima nas kolikšna je bila povprečna stopnja rasti te spremenljivke v obdobju 1990-1996. Povprečno stopnjo rasti izračunaj s pomočjo vrednosti verižnega indeksa in s pomočjo dejanske vrednosti spremenljivke.

VAJA 8 – MERE VARIABILNOSTI

Srednja vrednost je predstavnik vseh vrednosti spremenljivke, vendar je njena informativnost relativno majhna. Podatek, da se je z virusom COVID19 v decembru 2020 v povprečju okužilo 2000 ljudi, pove nekaj o srednji vrednosti števila okužb, nič pa o tem, kako je število okužb variiralo. Posledično je za dobro vrednotenje potreben še podatek o variabilnosti števila okuženih. Slednje opisujejo mere variabilnosti, ki vsebinsko dopolnjujejo mere sredine. Tako kot sredino spremenljivke lahko vrednotimo na različne načine, tudi variabilnost spremenljivke vrednotimo z različnih zornih kotov. Obravnavamo samo mere variabilnosti samo za številske spremenljivke. Ločimo absolutne (razmiki [variacijski in kvartilni] in odkloni [varianca in standardni odklon]) in relativne mere variabilnosti (koeficient variacije).

Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. Spremljamo vzorec slovenskih rek ($N=44$) s spremenljivko dolžina reke (v km). Izračunajte:
 - povprečno dolžino rek v vzorcu
 - odklone enot spremenljivke od aritmetične sredine
 - vsoto kvadratov odklonov
 - varianco
 - standardni odklon
 - koeficient variacije
2. Spremenljivka povprečna višina padavin (v mm) na izbranih meteoroloških postajah v Sloveniji je že razvrščena v frekvenčno porazdelitev z sedmimi razredi enake širine. Izračunaj naslednje kazalce:
 - aritmetično sredino
 - odklone po razredih
 - vsoto kvadratov odklonov
 - varianco
 - standardni odklon
 - koeficient variacije

VAJA 9 – NORMALNA PORAZDELITEV (KOLMOGOROV-SMIRNOV PREIZKUS)

Normalna porazdelitev sodi med najpomembnejše verjetnostne porazdelitve. Leto 1809 predstavlja pomemben mejnik pri razvoju zakonitosti, ki veljajo za normalno porazdelitev, ko je Carl Friedrich Gauss proučeval porazdelitev izmerkov neke količine ob tako imenovanih normalnih pogojih (ob slučajnih vplivih). Gostota verjetnosti normalne porazdelitve je definirana na celotni realni osi, je zvonaste oblike, je simetrična okoli aritmetične sredine, ima maksimum v vrednosti aritmetične sredine in se asimptotično bliža abscisni osi v obe smeri neskončnosti (pozitivno in negativno). V uporabi je predvsem najenostavnejša oblika normalne porazdelitve tj. standardizirana normalna porazdelitev $N(0,1)$. V splošnem velja, da lahko vsako normalno porazdelitev pretvorimo v standardizirano normalno porazdelitev kar olajša statistično sklepanje.

Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. Spremljamo vzorec meritev pojava mestnega toplotnega otoka ($N=100$) s spremenljivko temperatura zraka (v °C). Za omenjeno spremenljivko izračunajte naslednje kazalce s katerimi lahko sklepate o izpolnjevanju pogojev normalne porazdelitve:
 - aritmetično sredino
 - standardni odklon
 - skewness
 - kurtosis
2. Za spremenljivko temperatura zraka izračunajte gostoto verjetnosti normalne porazdelitve in kumulativno porazdelitveno funkcijo ter narišite gostoto verjetnosti normalne porazdelitve za obravnavani vzorec
3. Z-standardizirajte spremenljivko temperatura zraka, izračunajte gostoto verjetnosti normalne porazdelitve in kumulativno porazdelitveno funkcijo za Z-standardizirano spremenljivko ter narišite gostoto verjetnosti normalne porazdelitve za Z-standardizirani obravnavani vzorec
4. Izračunajte kolikšna je verjetnost če je vzorčna Z vrednost manjša ali enaka oziroma večja od 0, 1, 1.5, 2, 2.1, -1, -1.5, -2
5. Imamo vzorec slovenskih občin ($N=211$) z spremenljivko površina (km^2). Ugotovite ali gre za normalno porazdeljeno spremenljivko, uporabite Kolmogorov-Smirnov (KS) preizkus ($\alpha=0.05$). Zapišite ničelno domnevo KS preizkusa. V kolikor spremenljivka ni normalno porazdeljena jo ustrezno transformirajte. Nato ponovite testiranje domneve o normalni porazdelitvi ($\alpha=0.05$).

VAJA 10 – F PREIZKUS

Gre za parametrični statistični preizkus s katerim testiramo domnevo o enakosti varianc med dvema skupinama (lokalitetama) obravnavane normalno porazdeljene zvezne številske spremenljivke. Zagotovljeni morajo biti naslednji pogoji: (a) številska spremenljivka je zvezna, (b) je normalno porazdeljena (KS test v korist alternativne domneve), število enot spremenljivke v obeh skupinah (lokalitetah) je podobno.

Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. V MS Excel inštalirate dodatek za analizo podatkov (Datoteka → Možnosti → Dodatki → Orodja za analizo – VBA → Pojdi → Obkljukate vse dodatke (Orodja za analizo, Orodja za analizo – VBA, Orodja za valuto evro, Reševalnik) → Vredu → pod zavihkom Podatki v orodni vrstici se pojavi ikona Analiza podatkov
2. Imamo vzorec (N=30) povprečnih letnih temperatur zraka (v °C) za meteorološki postaji Maribor in Portorož. Preverite ali gre v obeh primeri za normalno porazdeljeno zvezno številsko spremenljivko (porabite Kolmogorov-Smirnov preizkus). Izračunajte kazalca kurtosis in skewness.
3. Izračunajte vzorčni varianci obravnavne spremenljivke za obe meteorološki postaji Maribor in Portorož.
4. Zapišite ničelno in alternativno domnevo F-preizkusa.
5. S pomočjo orodja Analiza podatkov izvedite ustrezni F-preizkus ($\alpha=0.05$) in označite potrjeno hipotezo.

VAJA 11 – T PREIZKUS

Gre za parametrični statistični preizkus s katerim testiramo domnevo o enakosti povprečij (aritmetičnih sredin) med dvema skupinama (lokalitetama) obravnavane normalno porazdeljene zvezne številske spremenljivke. Zagotovljeni morajo biti naslednji pogoji: (a) številska spremenljivka je zvezna, (b) je normalno porazdeljena (KS test v korist alternativne domneve), število enot spremenljivke v obeh skupinah (lokalitetah) je podobno, (d) testirana domneva o enakosti varianc (izveden F-preizkus). Točka d nam omogoča, da lahko izberemo ustrezno obliko T-preizkusa, saj se glede na potrditev ali zavrnitev domneve o enakosti vzorčnih varianc, računanje testne statistike in stopinj prostosti pri T-preizkusu značilno razlikuje.

Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. V MS Excel inštalirate dodatek za analizo podatkov (Datoteka → Možnosti → Dodatki → Orodja za analizo – VBA → Pojdi → Obkljukate vse dodatke (Orodja za analizo, Orodja za analizo – VBA, Orodja za valuto evro, Reševalnik) → Vredu → pod zavihkom Podatki v orodni vrstici se pojavi ikona Analiza podatkov
2. Imamo vzorec ($N=30$) povprečnih letnih temperatur zraka (v °C) za meteorološki postaji Maribor in Portorož. Izračunajte vzorčni varianci obravnavne spremenljivke za obe meteorološki postaji Maribor in Portorož. Zapišite ničelno in alternativno domnevo F-preizkusa. S pomočjo orodja Analiza podatkov izvedite ustrezni F-preizkus ($\alpha=0.05$) in označite potrjeno hipotezo.
3. Preverite ali obstajajo statistično značilne razlike v povprečni temperaturi zraka obeh meteoroloških postaj ($\alpha=0.05$). Zapišite ničelno in alternativno domnevo. Uporabite orodje Analiza podatkov in izberite ustrezno obliko T-preizkusa. Označite potrjeno hipotezo in zapišite ugotovitve.

VAJA 12 – ANOVA in KRUSKAL-WALLIS PREIZKUS

Pri analizi variance (ANOVA) gre za parametrični statistični preizkus s katerim testiramo domnevo o enakosti povprečij (aritmetičnih sredin) med več skupinami (lokalitetami) obravnavane normalno porazdeljene zvezne številske spremenljivke. Zagotovljeni morajo biti naslednji pogoji: (a) številska spremenljivka je zvezna, (b) je normalno porazdeljena, (c) število enot spremenljivke v vseh skupinah (lokalitetah) je podobno.

Pri Kruskal-Wallis preizkusu gre za ne-parametrični statistični preizkus s katerim testiramo domnevo o enakosti povprečij (aritmetičnih sredin) med več skupinami (lokalitetami) obravnavane ne-normalno porazdeljene zvezne ali ne-zvezne številske spremenljivke. Je ne-parametrični ekvivalent analizi variance. Uporabljamo ga, kadar obravnavana zvezna številska spremenljivka ne ustreza pogojem normalne porazdelitve in je ne moremo transformirati v korist pogojev normalne porazdelitve. Lahko ga uporabljamo tudi pri nezveznih številskih spremenljivkah (frekvencah).

Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. Obiščete spletno stran <https://www.real-statistics.com/free-download/real-statistics-resource-pack/> in sledite navodilom inštalacije paketa REAL-STATISTICS za vašo različico MS Excel
2. Ugotavljamo statistične razlike ($\alpha=0.05$) v povprečni višini padavin (v mm) med štirimi lokalitetami (Maribor, Ljubljana, Murska Sobota in Portorož). Gre za normalno porazdeljeno zvezno številsko spremenljivko. S pomočjo orodja Analiza podatkov izvedite ustrezno analizo variance (ANOVA) in zapišite ničelno ter alternativno domnevo. Označite katero domnevo sprejmemo glede na rezultate ANOVA.
3. Zaženite bližnjico do orodja REAL-STATISTICS (ctrl+m). Odprite zavihek Anova. Izberite možnost One Factor Anova. Označite polja podatkov, izberite možnost ANOVA in obkljukajte Tukey HSD pri okencu Anova Follow-up options. Greza za analizo primerjave parov, ki nam pove med katerimi skupinami so značilne razlike v povprečni vrednosti vhodne spremenljivke. Zapišete ničelno in alternativno domnevo. Glede na rezultate označite katera domneva je sprejeta in zapišite med katerimi meteorološkimi postajami obstajajo statistično značilne razlike v povprečni višini padavin ($\alpha=0.05$).
4. Na drugem listu so podatki o številu dni z močnim vetrom na enakih meteoroloških postajah. Tokrat gre za nezvezno spremenljivko. Predhodno vrednostim spremenljivke pripišemo range. Uporabimo orodje REAL-STATISTICS (ctrl+m) in izberemo zavihek Anova. Določi polje podatkov. Izberi Kruskal-Wallis preizkus in Dunn preizkus v rubriki Kruskal-Wallis follow-up options. Zapišete ničelno in alternativno domnevo. Glede na rezultate označite katera domneva je sprejeta in

zapišite med katerimi meteorološkim postajami obstajajo statistično značilne razlike v povprečni višini padavin ($\alpha=0.05$).

VAJA 13 – χ^2 PREIZKUS

Z hi-kvadrat preizkusom obravnavamo frekvenčne oz. kontingenčne tabele. Običajno v geografiji gre za testiranje zveze dveh kategoričnih spremenljivk (naklonski razred in raba tal). Pogosto ta statistični preizkus uporabljamo tudi na podatkih pridobljenih s pomočjo anketnih vprašalnikov kjer imamo opravka z nezveznimi, kategoričnimi spremenljivkami. Dodaten kazalec jakosti povezave dveh kategoričnih spremenljivk je korelacijski koeficient kontingence, ki lahko zaseda vrednosti na intervalu od 0 do 1. Tak primer sta popravljeni Pearsonov korelacijski koeficient kontingence (C_{pop}) in Cramerjev koeficient kontingence (V), ki sta pogosto dodatek k rezultatom hi-kvadrat preizkusa.

Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. Obiščete spletno stran <https://www.real-statistics.com/free-download/real-statistics-resource-pack/> in sledite navodilom inštalacije paketa REAL-STATISTICS za vašo različico MS Excel
2. Zanima nas povezava med tipom podlage in številom kmetij. Gre za kontingenčno tabelo 5x2. Izračunajte robne frekvence. Zapišite ničelno in alternativno domnevo hi-kvadrat preizkusa in stopnjo značilnosti. Odprite orodje REAL-STATISTICS (ctrl+m). Pomaknite se na zavihek Misc. Izberite možnost Chi-square Test for Independence. Vstavite polja z podatki (brez robnih frekvenc) skupaj z imeni stolpcev. Označite, da so imena stolpcev vključena v polje s podatki. Izberite Excel format. Označite, da gre za Input Format Categorical rows and columns. Doličite celico izpisa. Ostale nastavitve pustite po default-u. Rezultat sta dve tabeli. Prva prikazuje pričakovane frekvence po stolpcih, druga pa povzetek hi-kvadrat preskusa z pripadajočim Cramerjevim koeficientom kontingence. Glede na rezultate označite katera hipoteza je sprejeta.
3. Postopek v točki dva ponovite na podatkih o starostni strukturi dveh naselij in ugotovite ali obstajajo statistično značilne razlike v starostni strukturi obravnavanih naselij x in y. Zapišite ničelno in alternativno domnevo hi-kvadrat preizkusa in stopnjo značilnosti. Izračunajte pričakovane frekvence, testno statistiko in V. Označite katero hipotezo lahko sprejmemo glede na rezultat hi-kvadrat preizkusa.
4. Postopek v točki dva ponovite na podatkih o naklonu in rabi tal v občini Maribor in ugotovite ali obstaja statistično značilna zveza med rabo tal in naklonom na raziskovanem območju. Zapišite ničelno in alternativno domnevo hi-kvadrat preizkusa in stopnjo značilnosti. Izračunajte pričakovane frekvence, testno statistiko in V. Označite katero hipotezo lahko sprejmemo glede na rezultat hi-kvadrat preizkusa.

VAJA 14 – REGRESIJA (LINEARNI TREND)

Če ena spremenljivka vpliva na drugo, pravimo da sta regresijsko odvisni. Obe spremenljivki sta numerični. Odvisnost zapišemo v obliki funkcije $y = f(x)$. Mogočih je več vrst take odvisnosti, saj je tipov funkcij zelo veliko. Proučevali bomo le linearno odvisnost, ki jo lahko zapišemo z linearno funkcijo ali premico tipa $y = kx + n$. Odvisnost poznamo, če poznamo koeficienta k in n . Lahko bi matematični problem opredelili tudi tako: iz vzorca podatkov izračunamo koeficienta premice k in n tako, da se premica čim manj odklanja od točk (metoda najmanjših kvadratov). Pri korelaciji in drugih merah jakosti povezanosti spremenljivk običajno ugotavljamo le jakost zveze in njen predznak (+ ali -). Pri regresijah pa ugotavljamo smer vpliva, kaj je vzrok in kaj posledica, in ta vpliv kvantificiramo – govorimo o odvisnosti ene spremenljivke (Y) od druge ali več drugih (X_i).

Več vsebine najdete v knjigi Uporabna statistika (<https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>).

Naloga:

1. Obiščete spletno stran <https://www.real-statistics.com/free-download/real-statistics-resource-pack/> in sledite navodilom inštalacije paketa REAL-STATISTICS za vašo različico MS Excel
2. Zanima nas kako gostota zazidave vpliva na temperaturo zraka v urbanem sistemu (pojav mestnega toplotnega otoka). Izračunajte Pearsonov korelacijski koeficient. Narišite raztreseni diagram in ustrezno izberite odvisno in neodvisno spremenljivko. V grafikon dodajte enačbo premice in determinacijski koeficient (R^2).
3. Testirajte ali gre za značilno linearno povezanost obravnavanih spremenljivk. Uporabite orodje REAL-STATISTICS (ctrl+m). Izberite zavihek Reg in možnost Multiple Linear Regression. Vstavite polje z neodvisno spremenljivko. Vstavite polja z vrednostjo odvisne spremenljivke. Obkljukajte možnost Include constant term (intercept). Rubriki Options izberite možnosti Regression Analysis in Normality Test. Določite celico za izpis rezultatov (ostale parametre pustite po default-u) in pritisnite OK. Rezultati so prikazani v petih tabelah. OVERALL FIT prikazuje vrednosti Pearsonovega korelacijskega koeficienta, R^2 , prilagojeno R^2 vrednost, povprečno napako linearnega modela (Standard Error) in število podatkov. Druga tabela prikazuje kazalce za primerjavo modelov (AIC, AICc, SBC). Tretja tabela z naslovom ANOVA prikazuje statistično značilnost linearnega trenda. Četrta tabela prikazuje statistični povzetek vrednosti koeficientov k in n . Peta tabela z naslovom Shapiro-Wilk Test pa prikazuje pogoje normalne porazdelitve obeh spremenljivk. Na podlagi rezultatov oceni ali gre za statistično značilen linearen trend in ali gostota zazidave statistično značilno vpliva na temperaturo zraka v obravnavanem urbanem sistemu ter kolikšna je napaka v temperaturi zraka v kolikor predpostavljamo da na temperaturo zraka v mestu linearno vpliva le gostota zazidave. Kateri drugi dejavniki še vplivajo na pojav mestnega toplotnega otoka?

VIRI IN LITERATURA

- Košmelj K. (2007). Uporabna statistika (Druga dopolnjena izdaja). Elektronski vir: <https://repozitorij.uni-lj.si/Dokument.php?id=132497&lang=slv>
- STAT.SI → spletna baza podatkov Statističnega urada RS <https://pxweb.stat.si/SiStat/sl>
- STAGE → spletna baza prostorskih podatkov Statističnega urada RS <https://gis.stat.si/>
- EUROSTAT → spletna baza podatkov Evropske komisije <https://ec.europa.eu/eurostat/web/main/data/database>
- WORLD BANK → globalna spletna baza podatkov pod okriljem združenja petih inštitucij IBRD, IDA, IFC, MIGA in ICSID <https://www.worldbank.org/>
- ARSO → arhiv meritev <http://meteo.arso.gov.si/met/sl/app/webmet/#webmet==8Sdwx2bhR2cv0WZ0V2bvEGcw9ydlJWblR3LwVnaz9SYtVmYh9iclFGbt9SaulGdugXbsx3cs9mdl5WahxXYyNGapZXZ8tHZv1WYp5mOnMHbvZXZulWYnwCchJXYtVGdlJnOn0UQQdSf;>