

Univerzitet u Beogradu
Matematički fakultet

**Određivanje povezanosti deviznog kursa i
nivoa plata i penzija u Srbiji metodama
istraživanja podataka**
- Master rad -

Marijana Lazić

Beograd, 2013.

Univerzitet u Beogradu – Matematički fakultet

Master rad

Kandidat: Marijana Lazić

Naslov: Određivanje povezanosti deviznog kursa
i nivoa plata i penzija u Srbiji metodama
istraživanja podataka

Mentor: dr Nenad Mitić

Članovi komisije: dr Vesna Jevremović
dr Saša Malkov

Datum: Jun 2013.

SADRŽAJ

1.	Uvod	1
2.	Uvod u istraživanje podataka	2
2.1.	Šta je istraživanje podataka?	2
2.2.	Primeri primene istraživanja podataka	3
2.3.	Zadaci istraživanja podataka	4
2.4.	Postupak otkrivanja znanja	5
2.5.	Tipovi podataka	8
2.6.	Izbor metode u zavisnosti od tipa atributa	10
3.	Opis metoda	11
3.1.	Vremenske serije	11
3.1.1.	Predviđanje vremenskih serija	11
3.1.2.	Analiza vremenskih serija	12
3.1.2.1.	Određivanje trenda	14
3.1.2.2.	Izdvajanje sezonske komponente	16
3.2.	Korelacija	17
3.2.1.	Pirsonov koeficijent korelacije	20
3.2.2.	Spirmanova korelacija ranga	22
4.	Korišćeni softverski alati	24
4.1.	IBM SPSS Statistics	24
4.2.	EViews	24
5.	Prikupljanje i skladištenje podataka	25
6.	Upoznavanje sa podacima	28
6.1.	Devizni kurs	28
6.2.	Zarada	31
6.3.	Penzija	31

7.	Priprema podataka.....	33
7.1.	<i>Popunjavanje nedostajućih vrednosti.....</i>	33
7.2.	<i>Otkrivanje netipičnih podataka</i>	33
7.2.1.	Otkrivanje netipičnih podataka u atributu EUR_RSD	34
7.2.2.	Otkrivanje netipičnih podataka u atributu USD_RSD	36
7.2.3.	Otkrivanje netipičnih vrednosti u atributu neto_zarade	37
7.2.4.	Otkrivanje netipičnih vrednosti u atributu iznos_penzije	38
7.3.	<i>Transformacija podataka.....</i>	38
7.3.1.	Transformisanje dnevnih podataka u prosečne mesečne	38
7.3.2.	Transformisanje vrednosti atributa neto_zarade u normalno distribuirane podatke	39
8.	Karakteristike atributa.....	43
8.1.	<i>Karakteristike atributa EUR_RSD</i>	43
8.1.1.	Deskriptivna statistika	43
8.1.2.	Test normalnosti za atribut EUR_RSD	43
8.1.3.	Analiza trenda	44
8.1.4.	Analiza sezonske komponente	46
8.2.	<i>Karakteristike atributa USD_RSD</i>	47
8.2.1.	Deskriptivna statistika	47
8.2.2.	Test normalnosti za atribut USD_RSD	48
8.2.3.	Analiza trenda	49
8.2.4.	Analiza sezonske komponente	53
8.3.	<i>Karakteristike atributa neto_zarade</i>	53
8.3.1.	Deskriptivna statistika	54
8.3.2.	Test normalnosti za atribut neto_zarade	54
8.3.3.	Analiza trenda	55
8.3.4.	Analiza sezonske komponente	60
8.4.	<i>Karakteristike atributa iznos_penzije.....</i>	62
8.4.1.	Deskriptivna statistika	62
8.4.2.	Test normalnosti za atribut iznos_penzije	63
8.4.3.	Analiza trenda	64
8.4.4.	Analiza sezonske komponente	67
9.	Rezultati ispitivanja korelacije atributa	69
9.1.	<i>Analiza korelacije deviznog kursa EUR_RSD i prosečne neto zarade u periodu od 2003. do 2012. godine.....</i>	69

9.2.	<i>Analiza korelacije deviznog kursa EUR_RSD i prosečne penzije u periodu od 2003. do 2012. godine.....</i>	<i>70</i>
9.3.	<i>Analiza korelacije deviznog kursa USD_RSD i prosečne neto zarade u periodu od 2003. do 2012. godine.....</i>	<i>70</i>
9.4.	<i>Analiza korelacije deviznog kursa USD_RSD i prosečne neto zarade u periodu od 2003. – 2009. i u periodu od 2011. – 2012. godine</i>	<i>71</i>
9.5.	<i>Analiza korelacije deviznog kursa USD_RSD i prosečne penzije u periodu od 2003. do 2012. godine.....</i>	<i>72</i>
9.6.	<i>Analiza korelacije deviznog kursa USD_RSD i prosečne penzije u periodu od 2003. – 2009. i u periodu od 2011. – 2012. godine</i>	<i>73</i>
9.7.	<i>Analiza korelacije prosečne neto zarade i broja zaposlenih</i>	<i>73</i>
10.	Zaključak	75
	Dodaci.....	78
10.1.	<i>Dodatak A: Primer određivanja pokretnog proseka.....</i>	<i>78</i>
10.2.	<i>Dodatak B: Primer određivanja sezonskog indeksa.....</i>	<i>81</i>
10.3.	<i>Dodatak C: Primer desezoniranja</i>	<i>84</i>
10.4.	<i>Dodatak D: Normalna raspodela.....</i>	<i>87</i>
10.5.	<i>Dodatak E: Primer Spirmanove korelacije ranga.....</i>	<i>94</i>
10.6.	<i>Dodatak F: Program XLSManipulation.....</i>	<i>96</i>
10.7.	<i>Dodatak G: Program HTMLManipulation.....</i>	<i>98</i>
11.	Literatura	102

1. Uvod

U priloženom radu su predstavljeni rezultati ispitivanja povezanosti promene deviznih kurseva evra i dolara sa jedne strane i promene prosečnih penzija i zarada zaposlenih u Republici Srbiji u periodu od 2003. do 2012. godine sa druge strane. Ove povezanosti su analizirane postupkom otkrivanja znanja u podacima (eng. *Knowledge Discovery in Database* ili skraćeno KDD) i softverskim alatima za istraživanje podataka. Kao glavna metoda za ispitivanje da li su dva atributa povezana, primenjena je Spirmanova korelacija ranga. Ova metoda je izabrana jer karakteristike datih atributa ne ispunjavaju preduslov za primenu drugih metoda. Značaj analiza predstavljenih u ovom radu se ogleda u mogućnosti predviđanja ponašanja deviznih kurseva i zarade.

Sažetak:

- U drugom poglavlju je ukratko predstavljen proces istraživanja podataka (šta je, kojim zadacima se bavi, koji su koraci, i primeri praktične primene).
- U trećem poglavlju su predstavljene metode koje će biti korišćene za klasičnu analizu vremenskih serija, kao i metode za ispitivanje povezanosti dva atributa.
- U četvrtom poglavlju su opisani softverski alati koje su korišćeni tokom rada.
- U petom poglavlju je predstavljen tip i struktura skladišta podataka, kao i izvori iz kojih su podaci preuzeti.
- U šestom poglavlju čitalac se upoznaje sa osnovnim informacijama o podacima koje će u radu biti analizirani - deviznom kursu, zaradama i penzijama.
- U sedmom poglavlju se podaci obrađuju i pripremaju za predstojeću analizu. Atributi se čiste od netipičnih podataka i nedostajućih vrednosti.
- U osmom poglavlju se utvrđuju karakteristike atributa u cilju što boljeg izbora metode za ispitivanje povezanosti. Izvršena je klasična analiza vremenskih serija, kao i testovi normalnosti.
- U devetom poglavlju predstavljeni su rezultati primene metoda za ispitivanje korelacija.
- U dodacima su dati primeri primene svih korišćenih metoda kako bi dobijeni rezultati svakom čitaocu, bez obzira na njegovo predznanje, bili jasni.

2. Uvod u istraživanje podataka

Količina podataka koja se svakodnevno skladišti u raznim bazama podataka je drastično porasla u poslednjih nekoliko godina. Danas, takve baze podataka mogu da imaju i po nekoliko desetina terabajta. Poreklo ove količine podataka može biti različito. To mogu biti podaci prikupljeni sa računara iz prodavnica, bankovnih transakcija, iz telefonskih anketa, itd. U ovim neobrađenim podacima skrivene su informacije koje su od velikog značaja za različite organizacije (npr. organizacije koje se bave naučnim istraživanjima ili kompanije koje žele da ispituju i unaprede svoje poslovanje).

Podatak → Informacija

Podatak se može definisati kao re-interpretabilna reprezentacija činjenice ili ideje na formalan način pogodan za komunikaciju, interpretaciju ili obradu. Operacije nad podacima mogu da vrše ljudi ili mašine [1]. **Informacija** je dodela značenja podatku.

S obzirom na to da zbog količine podataka postupak uočavanja bitnih informacija tradicionalnim (ručnim) tehnikama nije efikasan, mora se pronaći automatizovani pandan ovim tehnikama, koji će davati bar iste rezultate kao i tradicionalne tehnike, ali će za razliku od njih, moći jednostavno i bez grešaka da se primeni na velikim količinama podataka. Odgovor se nalazi u procesu istraživanja podataka.

2.1. Šta je istraživanje podataka?

Istraživanje podataka (eng. *Data mining*) je proces otkrivanja korisnih informacija u bazama podataka, korišćenjem računarskih alata [2].

Kako se javila ideja za ovakav naziv?

Istraživanje podataka (u direktnom prevodu sa engleskog na srpski jezik – rudarenje podataka) ima naziv koji figurativno opisuje zadatak kojim se bavi. Naime, postupkom istraživanja podataka se „kopa“ (istražuje) po gomili sirovih podataka u potrazi za rudom tj. informacijama koje su zlata vredne (slika 1) ¹.

¹ Slika je preuzeta sa veb sajta: <http://kenmillergroup.com>



Slika 1. Slika koja figurativno opisuje zadatak istraživanja podataka

Zahvaljujući alatima za istraživanje podataka moguće je doći do značajnih informacija i odgovora na raznovrsna pitanja analizom prethodnih zapisa o posmatranoj pojavi. To su npr. informacija o tome kakvo će tržište biti za dva dana, koliki će kurs biti, šta je uticalo na kupovinu nekog proizvoda, itd. Odgovori na ova i slična pitanja dobijaju se analizom baze podataka prethodnih vrednosti posmatrane pojave sa raznih stanovišta, primenom različitih matematičkih algoritama i pronalaženjem veza i odnosa između naizgled nepovezanih informacija.

2.2. Primeri primene istraživanja podataka

Primer primene u trgovini

Zahvaljujući fiskalnim računima lanac trgovinskih prodavnica „X“, ima svakog dana na raspolaganju znatnu količinu podataka o proizvodima koji su tokom tog dana kupljeni u svim njihovim prodavnicama. Rukovodstvo ovog lanca želi da na osnovu tih podataka otkrije određene šablone ponašanja kupaca, kako bi se u skladu sa tim napravila marketinška i investiciona strategija. Jasno je da bi ručno pretraživanje ove baze podataka u potrazi za bilo kakvom smislenom informacijom predstavljalo „nemoguću misiju“, te se zato koristi istraživanje podataka kako bi se upotrebom računara i statističkih metoda brzo došlo do odgovora. Računar, u kombinaciji sa odgovarajućom tehnikom istraživanja podataka, može da obradi sve račune i pronađe u njima određene šablone i veze. Tako se, na primer, računarskom obradom podataka o kupljenim proizvodima (pribavljenim sa računa) može uočiti šablon da skoro svaki kupac koji kupi pivo, kupi i kikiriki. I u tom slučaju, rukovodstvo lanca trgovinskih prodavnica može doneti odluku da se kikiriki premesti pored piva.

Primer primene u političkoj kampanji

Istraživanje podataka je imalo veliku ulogu tokom planiranja kampanje za predsedničke izbore 2012. godine u SAD. Sakupljani su podaci o računarski nezasnovanim aktivnostima glasača (kao što su učestvovanje u humanitarnim akcijama putem mobilnih telefona, podaci o transakcijama na kreditnim karticama, informacije o polu, rasi, seksualnoj opredeljenosti, konzumiranje alkohola, duvana, da li imaju decu, itd.) i podaci o Internet zasnovanim aktivnostima (kao što su omiljeni veb sajtovi, „lajkovi“ na *Facebooku*, prijatelji na društvenim mrežama, grupe u koje su učlanjeni, reklame koje su ih zainteresovale, šta su pretraživali u pretraživačima²). Procesom istraživanja podataka dobijene su informacije na osnovu kojih su marketinški timovi predsedničkih kandidata stekli jasniju sliku o postojećim glasačkim grupama, a onda u smeru brojčano većih grupa usmeravali svoju kampanju. Npr., ako se ustanovi da 80% glasačkog tela konzumira alkohol, u predizbornoj kampanji će se sigurno zalagati za donošenje zakona o prodaji alkohola posle 22h.

Primer primene istraživanja podataka u bankama

Banka za svakog klijenta vodi evidenciju o računima, transakcijama, kreditnim obavezama, bračnom stanju, itd. Na osnovu ovih podataka banka može odrediti potencijalne klijente koji po proceni nekog algoritma istraživanja podataka neće vratiti kredit, i tako sprečiti novčani gubitak [3].

Postoji veliki broj različitih primera primene istraživanja podataka u skoro svakoj oblasti – npr. mogu se istraživati indikatori koji izazivaju neku bolest; može se na osnovu ključnih reči u poruci napraviti *spam filter* koji određuje da li je primljena elektronska pošta neželjena reklama, itd.

2.3. Zadaci istraživanja podataka

Gruba podela zadataka kojima se istraživanje podataka bavi bi, prema [2], bila na:

- **Prediktivne zadatke** – kod ovog tipa zadataka cilj je automatizovano predviđanje nove vrednosti atributa na osnovu već postojećih podataka koji se nalaze u bazi. Rezultat je nova/e vrednost/i, za razliku od deskriptivnog tipa zadataka gde je rezultat (ne)potvrđena hipoteza. Prediktivni zadaci pomažu organizacijama u donošenju boljih odluka, sa ciljem da uštede novac i naprave bolje planove za budućnost. Na primer, predviđanje rasta ili pada deviznog kursa za naredni dan ili nedelju može pomoći nekome da profitira prodajom ili kupovinom deviza na deviznom tržištu.
- **Deskriptivne zadatke** – kod ovog tipa zadataka cilj je pronaći neku pravilnost, identifikovati neke šablone i obrasce (korelacije, trendove, anomalije) u podacima skladištenim u bazi. Kod ovih tipova zadataka najpre se postavi hipoteza, koja se nakon istraživanja odbacuje ili potvrđuje u

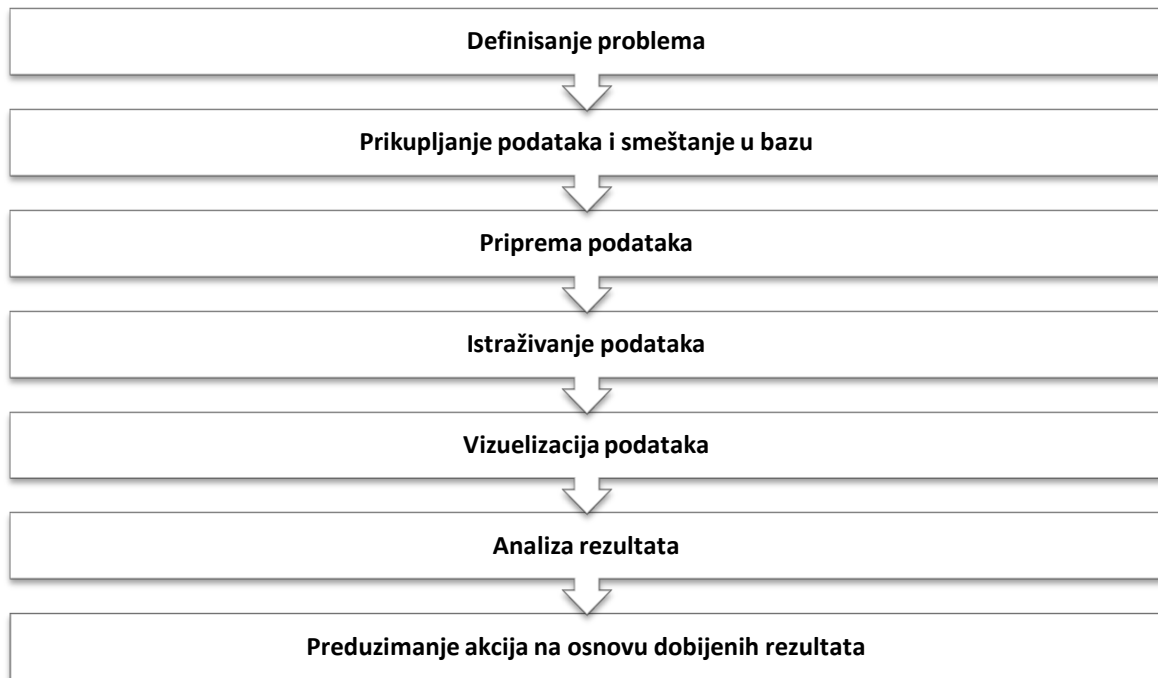
² Google omogućava pregled podataka koji se čuvaju o svakom registrovanom korisniku njegovih usluga. Čuva se šta je pretraživano, koji kontakt je najviše kontaktiran, od kog kontakta je primljeno najviše pošiljke, itd. Ove informacije se mogu pronaći u „Izveštaju o aktivnostima“ na veb adresi: <https://www.google.com/settings/activity/manage>

zavisnosti od dobijenih rezultata analize. Npr., može se doći do zaključka da kupac koji kupi belo vino, obično kupi i mineralnu vodu.

2.4. Postupak otkrivanja znanja

Otkrivanje znanja i Istraživanje podataka

Termin otkrivanje znanja se često upotrebljava kao sinonim za istraživanje podataka, iako to nije u potpunosti tačno. Naime, istraživanje podataka je zapravo samo deo procesa otkrivanja znanja. Korak istraživanja podataka jeste bitan, ali bez kompletiranja svih dole-navedenih faza otkrivanja znanja, istraživanje podataka može da proizvede netačne i beznačajne obrasce i informacije. Na slici 2 su prikazani koraci u postupku otkrivanja znanja.



Slika 2. Koraci otkrivanja znanja

Koraci otkrivanja znanja su:

- 1) Definisanje problema i ciljeva istraživanja
Ovo je početna faza u kojoj se vrši identifikovanje i definisanje cilja istraživanja ili problema koji treba da se reši, definisanje željenih rezultata, kao i detaljno upoznavanje sa područjem koje se istražuje.
- 2) Prikupljanje podataka i smeštanje u bazu

U ovoj fazi se vrši pribavljanje podataka neophodnih za analizu i njihovo smeštanje u bazu podataka. U zavisnosti od složenosti i količine podataka, kao i softvera koji će se koristiti za analizu, pažljivo se bira tip skladišta. Najčešće se koriste relacione baze podataka.

3) Priprema

Pre vršenja bilo kakve analize podataka potrebno je izvršiti pripremu podataka kako bi se dobilo nešto što je istražive prirode, jer najčešće nije moguće direktno korišćenje podataka iz baze. Nad podacima iz baze je poželjno uraditi sledeće zadatke:

Čišćenje podataka. Kako bi se dobili ispravni rezultati bazu je neophodno očistiti od duplih, nepotpunih, nedostajućih i netipičnih podataka.

- Netipični podaci (eng. *outlier*) su prisutni u gotovo svakoj bazi podataka. To su vrednosti koje se izrazito razlikuju od drugih vrednosti iz baze (npr. osoba koja je visoka 2 m). Dok se ovakve vrednosti ne smatraju greškom i u nekim analizama njihovo postojanje je poželjno, postoje i netipični podaci koji predstavljaju greške nastale pri unosu podataka i takve podatke treba ukloniti iz baze, jer mogu znatno da izmene rezultate analize.
- Ako postoje nedostajuće (nepotpune) vrednosti potrebno je utvrditi da li je moguće te podatke izračunati na osnovu nekih drugih vrednosti. Npr. ako nedostaje godina rođenja ili pol osobe, taj podatak se veoma jednostavno može dobiti na osnovu jedinstvenog matičnog broja građana. Ukoliko ne postoji način da se dobiju ti podaci onda se oni isključuju iz analize.

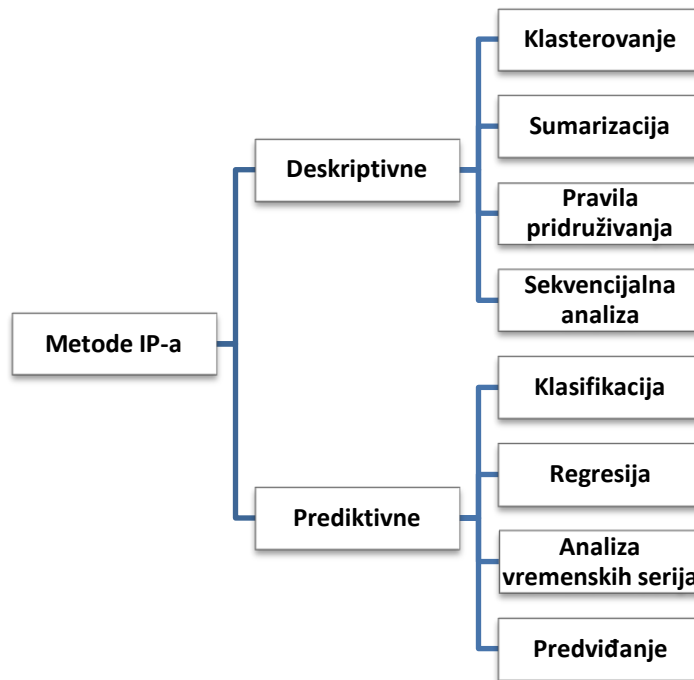
Transformacija vrednosti. Nad nekim podacima je neophodno izvršiti transformaciju da bi se nad njima mogli primeniti određeni algoritmi, npr. transformisanje kontinualnih u diskretne attribute (diskretizacija), logaritamska ili kvadratna transformacija kako bi se dobili normalno raspoređeni podaci, izračunavanje mesečnih proseka, itd.

Faze priprema i prikupljanje podataka su najzahtevnije faze otkrivanja znanja, pa se uglavnom 80% vremena troši upravo na ove dve faze. Pre prelaska na naredni korak, primenu algoritama za istraživanje podataka, veoma je značajno da se do ovog koraka podaci dobro istraže i shvate, jer dobro poznavanje podataka olakšava identifikovanje algoritama koji će se primeniti, kao i tumačenje dobijenih rezultata.

4) Istraživanja podataka

Nakon što se dobiju sređeni podaci, nastupa faza istraživanja podataka. Istraživanje podataka nije samostalna tehnika, ili tehnologija, nego skup srodnih metoda usmerenih ka pronalaženju i automatskom otkrivanju šablona, sličnosti, promena, anomalija i drugih karakterističnih struktura iz podataka. Postoje različiti algoritmi pomoću kojih se podaci mogu istraživati. Izbor algoritma zavisi od prirode podataka nad kojima se primenjuje, ali i od tipa rezultata koje želimo da dobijemo. Metode istraživanja podataka se mogu svrstati u dve kategorije, u zavisnosti od tipa

zadatka kojim se bave: deskriptivne i prediktivne. Na slici 3 je prikazana podela metoda istraživanja podataka [3]:



Slika 3. Metode istraživanja podataka

Metode otkrivanja znanja koje se bave deskriptivnim zadacima su klasterovanje, sumarizacija, pronalaženje pravila pridruživanja i sekvencijalna analiza.

- Klasterovanje se koristi za grupisanje podataka po sličnosti u klastere čiji broj nije unapred određen. Klasterovanje se zove još i učenje bez nadzora. Na primer, grupisanje kupaca po starosti, zanimanju, zaradi.
- Sumarizacija preslikava podatke u podskupove sa pridruženim opisima. Na taj način se izdvajaju reprezentativne informacije iz baze. Na primer, poređenje kvaliteta filmova po oceni kritičara na www.IMDb.com ili po broju dobijenih nagrada.
- Pravila pridruživanja se koriste za otkrivanje šablona koji opisuju međusobnu povezanost atributa. Na primer, promocije proizvoda - sniziti cenu mleka, ali povećati cenu hleba.
- Sekvencijalna analiza se koristi za pronalaženje šablona u podacima koji zavise od redosleda pojavljivanja. Na primer, sekvencijalna analiza DNK.

Metode otkrivanja znanja koje se bave prediktivnim zadacima su klasifikacija, regresija, analiza vremenskih serija i predviđanje.

- Klasifikacija se koristi za predviđanje osobina ili kategorija. Npr. da li će student završiti fakultet ili ne, da li će klijent vratiti dug ili ne, itd. Klasifikacija se zove još i učenje pod nadzorom jer su klase određene pre ispitivanja podataka.
- Regresija se koristi za predviđanje vrednosti promenljive čiji je domen skup realnih brojeva, na osnovu vrednosti drugih promenljivih. Na primer, predviđanje potrošnje struje u zavisnosti od doba dana i godišnjeg doba.
- Analiza vremenskih serija se koristi za istraživanje vrednosti atributa kroz vreme tj. za ispitivanje ponašanja atributa kroz vreme.
- Metode za predviđanje vrednosti promenljivih se koriste za predviđanje numeričkih vrednosti. Npr. predviđanje ocene koju će student dobiti na osnovu njegovih ostalih ocena.

Sve metode su uključene u različite softverske alate kao što su SPSS, Eviews, IBM Infosphere, R, Weka, SAS, itd.

5) Vizuelizacija podataka

Vizuelizacija podataka predstavlja prikazivanje podataka u nekom grafičkom formatu i najčešće za cilj ima prepoznavanje nekih karakteristika podataka (npr. linearnost, raspodelu, itd.). I na osnovu vizuelizacije se može doći do značajnih zaključaka, relacija, šablona i identifikovanja netipičnih podataka. Da bi tumačenje bilo ispravno, neophodno je odabrati odgovarajuću tehniku vizuelizacije. Taj izbor zavisi od prirode i tipa podataka, ali postoje i generičke tehnike koje se mogu koristiti bez obzira na tip podataka. Neke od tehnika vizuelizacije su: histogrami, kućice, grafik disperzije, itd.

6) Analiza rezultata

Nakon što se izabere i primeni odgovarajući algoritam dobijaju se rezultati, koje je potrebno protumačiti i proveriti njihovu istinitost i značaj.

7) Preduzimanje akcija na osnovu dobijenih rezultata

Na osnovu dobijenih rezultata, organizacija/kompanija koja je vršila istraživanje, donosi odgovarajuću odluku koja se tiče njenog daljeg rada.

2.5. Tipovi podataka

Znati tipove podataka koji se analiziraju je od velike važnosti, jer od toga zavisi izbor tehnike koja će se koristiti za analizu. Loš izbor algoritma vodi do loših rezultata. Postoji nekoliko vrsta podele podataka:

- a) U zavisnosti od načina izražavanja vrednosti, podatke možemo razvrstati na kontinualne (zovu se još i kvantitativni) i kategoričke (zovu se još i kvalitativni, diskretni). Kontinualni atributi su oni koji imaju neprebrojivo mnogo vrednosti, npr. visina, težina, itd. Diskretni atributi su oni koji imaju najviše prebrojivo mnogo vrednosti. Primer diskretnih atributa su binarni atributi koji imaju samo vrednosti 0 i 1.

b) U zavisnosti od načina merenja, podaci se mogu podeliti na [2]:

- nominalne,
- ordinalne,
- intervalne,
- razmerne.

Prva dva tipa pripadaju kategoričkim tipovima, a druga dva kontinualnim.

Nominalni atributi

Nominalni atributi su atributi čije se vrednosti mogu svrstati u bar dve različite kategorije, istih važnosti, koje su nastale na osnovu neke kvalitativne osobine. Ovi atributi se ne mogu urediti po nekom redosledu, niti postoji kriterijum po kom bi se moglo utvrditi da li je neka vrednost tog atributa veća ili manja od neke druge vrednosti. Na primer:

- atribut *tip računa* može imati 3 kategorije: standardni, biznis, studentski.
- atribut *pol* može imati 2 kategorije: muški, ženski

Ordinalni atributi

Ordinalni atributi su atributi koji se isto kao i nominalni mogu svrstati u dve ili više kategorije, samo što je u ovom slučaju klasifikacija urađena po određenom redosledu (eng. *rank*). Na primer, ocene u školi su ordinalni atributi: 1 – nedovoljan; 2 – dovoljan; 3 – dobar; 4 – vrlo dobar; 5 – odličan. Nad vrednostima ordinalnih atributa se ne mogu primenjivati aritmetičke operacije. Vrednosti ovih atributa mogu biti numeričke, ali nema nikakvog smisla nad njima raditi sabiranje, oduzimanje, množenje ili deljenje. Npr. ništa se ne dobija oduzimanjem ocene 4 od ocene 2 - takva operacija nema nikakvog smisla.

Intervalni atributi

Intervalni atribut je atribut kod kojeg razlike između vrednosti, koje taj atribut ima, imaju smisao. Osobine i operacije čija primena ima smisla nad ovakvim podacima su: operacije sabiranja, oduzimanja, klasifikacije i rangiranja. Ovi atributi nemaju prirodnu nulu. Primeri intervalnog atributa je temperatura merena Celzijusovom skalom.

Razmerni atributi

Razmerni atributi su intervalni atributi kod kojih postoji prirodna nula. Na primer, ako je Celzijusovom skalom izračunato da temperatura ima vrednost 0°C, to ne znači da tada nema temperature. Ali, ako se temperatura meri Kelvinovom skalom i ima vrednost 0K to zaista znači da nema temperature. Sa vrednostima razmernih atributa dozvoljeno je vršiti poređenje mera i poređenje intervala, kao i aritmetičke operacije sabiranja, oduzimanja, množenja i deljenja. Primeri razmernih atributa su visina, težina, starost, broj nečega, itd.

2.6. Izbor metode u zavisnosti od tipa atributa

Kao što je već rečeno pre nego što se krene sa daljim radom veoma je bitno odrediti tip svih atributa uključenih u analizu, jer se u zavisnosti od tipa određuje i metoda koja će se koristiti za analizu. Pošto se u radu ispituje povezanost deviznog kursa i zarada, bitno je razmotriti koje se metode mogu koristiti za ispitivanje ove veze. U tabeli 1 su predstavljene metode za ispitivanje povezanosti atributa [9].

Cilj istraživanja	Karakteristike procesa merenja		
	Kvantitativno (iz normalno raspoređene populacije)	Rangovi, skorovi ili kvantitativno (iz kontinuirano ali ne-normalno raspodeljene populacije)	Kategorijalni – binomni (dva moguća ishoda)
Kvantifikacija povezanosti dve varijable	Pirsonova linearna korelacija	Spirmanova korelacija ranga	Koeficijent kointegracije
Predviđanje vrednosti jedne varijable na osnovu izmerenih vrednosti druge	Linearna ili nelinearna regresija	Neparametarska regresija	Jednostavna (univarijantna) logistička regresija
Predviđanje vrednosti jedne varijable na osnovu izmerenih vrednosti dve ili više varijabli	Višestruka linearna ili nelinearna regresija		Višestruka logistička regresija

Tabela 1. Metode za ispitivanje povezanosti atributa

Pošto su devizni kurs i zarada kontinualni atributi sa vrednostima koje nemaju normalnu raspodelu kao osnovni metod za ispitivanje veze između njih odabrana je Spirmanova korelacija ranga. Način ispitivanja uslova koji su uticali na odabir ove metode će biti kasnije detaljnije opisan.

3. Opis metoda

3.1. Vremenske serije

Svi podaci korišćeni u toku rada su predstavljeni u vidu vremenskih serija. Pre primene glavne metode – korelacije, neophodno je analizirati i pripremiti ove vremenske serije. U nastavku će biti opisane metode koje će u ove svrhe biti korišćene.

Šta je vremenska serija?

Vremenska serija je skup numeričkih vrednosti atributa X , dužine t : $X = x_1, x_2, \dots, x_t$ poređanih tako da reprezentuju promene atributa X u vremenu t [7]. Jednostavnije rečeno, vremenska serija je niz vrednosti dobijenih posmatranjem određene pojave kroz vreme.

Podaci u vremenskoj seriji se uvek odnose na prošle vremenske periode. Intervali merenja određene pojave mogu biti sati, dani, nedelje, meseci, kvartali, godine, itd. Za pojave koje slabo osciliraju tj. koje se sporo menjaju (na primer broj stanovnika) koriste se godišnji intervali kao vremenske jedinice, dok se za pojave koje se češće menjaju (kao na primer cene na malo) koriste mesečni ili dnevni intervali kao vremenske jedinice.

Primeri vremenskih serija su:

- posete veb sajtu beležene na sat vremena,
- prodaja proizvoda na mesečnom nivou,
- dnevna vremenska prognoza,
- itd.

Istraživanje vremenskih serija

Vremenske serije možemo koristiti za dva tipa istraživačkih zadataka: analizu (određivanje trenda, sezone, ciklusa, itd.) i predviđanje novih vrednosti.

3.1.1. Predviđanje vremenskih serija

Predviđanje (eng. *forecasting*) podrazumeva analizu istorijskih podataka date pojave (trend, sezonsku komponentu) i projektovanje istih u budućnost pomoću nekog matematičkog modela, kako bi se mogle prognozirati buduće vrednosti posmatrane pojave, pošto se pretpostavlja da će pojava koja se ispituje nastaviti da se ponaša slično kao što je u bliskoj prošlosti.

Postoji nekoliko metoda za predviđanje novih vrednosti:

- Eksponencijalno poravnanje,

- Boks–Dženkins ARIMA,
- Autoregresija,
- itd.

3.1.2. Analiza vremenskih serija

Analiza vremenskih serija ima za cilj da ustanovi dinamiku i svojstva određene pojave u određenom vremenskom intervalu, kao i pronalaženje određenih događaja i faktora koji su znatno promenili ponašanje serije tj. pojave koja se prati. Na primer, uzmimo ispitivanje nedeljnog broja saobraćajnih nesreća. Pri analizi se primećuje da je za 2002. godinu bilo u proseku 256 nesreća nedeljno, a od 01.04.2003. taj broj je smanjen na 156. Nakon utvrđivanja kada je došlo do znatne promene toka vremenske serije, sledeći zadatak je bio utvrđivanje razloga nastanka te promene. Istraživanjem se došlo do informacije da je 25.03.2003. stupio novi Zakon o saobraćaju, što je dovelo do ovakvog naglog pada broja saobraćajnih nesreća.

Dekompozicija vremenskih serija

U ovom radu će biti primenjena klasična metoda statističke analize vremenskih serija - metoda dekompozicije, koja polazi od pretpostavke da je vremenska serija komponovana na aditivan ili multiplikativan način od četiri komponente: trenda, ciklične, sezonske i slučajne komponente [7].

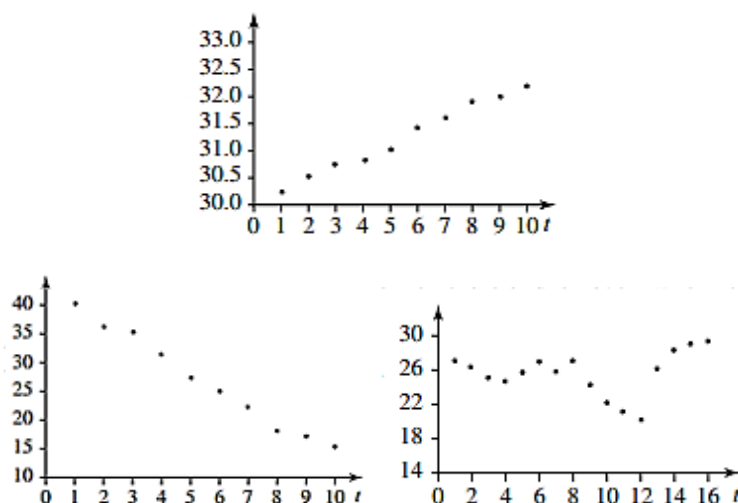
Glavna prednost metode dekompozicije je njegoa razumljivost, lakoća primene i jednostavnost interpretacije. Najveći nedostatak je to što ne postoji teorijski model na osnovu kojeg možemo izvršiti statističko vrednovanje dobijenih rezultata dekompozicije.

U nastavku će biti opisane komponente vremenske serije, kao i modeli na osnovu kojih se te komponente mogu komponovati.

Komponente vremenskih serija

Vremenske serije mogu imati četiri komponente [7]:

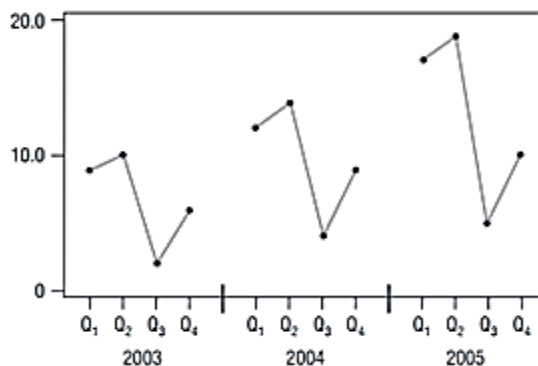
- 1) **Trend** predstavlja dugoročnu razvojnu tendenciju pojave koja se prati. Vrednosti posmatranog vremenskog niza mogu imati dugoročnu tendenciju rasta ili pada, pa se u skladu sa tim razlikuje rastući i opadajući trend. Pored ova dva, postoji i bočni trend koji se odnosi na vremensku seriju bez izraženog trenda. Na slici 4 su prikazani izgledi svih tipova trenda.



Slika 4. Rastući, opadajući i bočni (neutralni) trend

- 2) **Sezonska komponenta** predstavlja promene, nad pojavom koja se ispituje, nastale usled delovanja nekog faktora u isto vreme u toku godine i po istom šablonu (Slika 5). Da bi se analizirala ova komponenta, podaci moraju biti predstavljeni u mesečnom ili kvartalnom intervalu. Primer prodaje sladoleda: prodaja sladoleda je najveća tokom letnje sezone u junu, julu i avgustu svake godine.

Za vremensku seriju kažemo da je desezonirana, ako iz nje uklonimo sezonsku komponentu.



Slika 5. Izgled sezonske komponente na grafiku

- 3) **Ciklična komponenta** predstavlja varijacije u vremenskoj seriji koje se ponavljaju najčešće u nejednakim vremenskim razmacima, dužim od jedne godine, i sa nejednakom dužinom trajanja i intenzitetom. Postojanje ciklične komponente nije obavezno kod svake vremenske serije.
- 4) **Slučajna komponenta** predstavlja promene nastale usled uticaja slučajnih spoljnih faktora koje nije moguće predvideti. Svaka vremenska serija ima slučajnu komponentu jer se radi o statističkim podacima čija vrednost nije određena determinističkim procesima. Za primer prodaje

sladoleda slučajna komponenta bi bila kišovit dan u sezoni. Za vreme kišovitih dana u sezoni prodaje sladoleda kupovina je znatno manja.

Navedene komponente nemaju vrednost samo sa stanovišta istorijske analize prethodnih događaja, već i za otkrivanje budućih promena. Na primer, u slučaju privrede one omogućavaju preduzimanje pravovremenih mera u cilju stabilizovanja budućih privrednih kretanja.

Modeli dekompozicije vremenskih serija

Tipovi modela dekompozicije vremenskih serija su [15]:

- Aditivni
- Multiplikativni
- Mešoviti (kombinovani) model

Svaka komponenta posmatrane pojave u vremenskoj seriji se može izračunati pomoću ovih modela.

Prema multiplikativnom modelu vremenska serija je jednaka proizvodu svojih osnovnih komponenti [15]:

$$Trend * Ciklična * Sezonska * Slučajna$$

Prema aditivnom modelu vremenska serija je jednaka zbiru svojih osnovnih komponenti:

$$Trend + Ciklična + Sezonska + Slučajna$$

Pošto postoji mogućnost da neka od ovih komponenti ne postoji u vremenskoj seriji, tada je ta komponenta jednaka 0 u aditivnom modelu, a 1 u multiplikativnom. Jedino je slučajna komponenta uvek prisutna u svakoj vremenskoj seriji.

Prema mešovitom modelu, koji se veoma retko koristi, vremenska serija je jednaka zbiru proizvoda osnovnih komponenti, na primer:

$$Trend * Ciklična * Sezonska + Slučajna$$

3.1.2.1. Određivanje trenda

Trend oslikava dugoročnu tendenciju promene pravca posmatrane vrednosti kroz vreme. U većini slučajeva trend vremenske serije predstavlja komponentu od najvećeg značaja s obzirom na to da opisuje ponašanje date pojave u dužem vremenskom periodu. Trend u nekim slučajevima nije očigledan, pa se zbog toga mora primeniti odgovarajuća metoda za određivanje trenda.

Trend se može odrediti nekom od sledećih metoda [15]:

- metodom najmanjih kvadrata,
- metodom pokretnih proseka (pokretnih sredina),
- metodom apsolutne razlike,

- metodom Hold-Vintersa,
- metodom standardne greške,
- itd.

Detaljnije će biti opisana samo metoda pokretnih sredina, koja se koristi kada nije izražena ciklična komponenta, što odgovara podacima korišćenim u ovom radu.

Metoda pokretnih proseka

Jedna od najjednostavnijih metoda određivanja trenda, u slučaju da su vrednosti X_t date za ekvidistantne vrednosti promenljive t , je metoda pokretnih proseka. Metoda pokretnih proseka se najčešće koristi za određivanje trenda u slučaju da ciklična komponenta ne postoji, ili za određivanje proizvoda trenda i ciklične komponente u slučaju da ciklična komponenta postoji. Pokretni prosek predstavlja konstrukciju nove vremenske serije u kojoj je svaki originalni podatak vremenske serije zamenjen aritmetičkom sredinom tog podatka, određenog broja njegovih prethodnika i isto toliko sledbenika.

(2*k+1)-člani pokretni prosek (Neparni pokretni proseci)

Pokretni prosek ($\bar{x}_t$) za vremensku seriju $x_1, x_2, \dots, x_n$ sa neparnim brojem podataka se računa preko formule [8]:

$$\bar{x}_t = \frac{1}{2*k+1} * \sum_{j=-k}^k x_{t+j}, \quad i = k+1, k+2, \dots, n-k \quad (1)$$

U ovom slučaju se ne može izračunati pokretni prosek prvih i poslednjih k podataka.

(2*k)-člani pokretni prosek (Parni pokretni proseci)

Pokretni prosek ($\bar{x}_t$) za vremensku seriju $x_1, x_2, \dots, x_n$ sa parnim brojem podataka se računa preko formule [8]:

$$\bar{x}_t = \frac{x_{t-k} + \sum_{j=k+1}^{2k-1} 2x_{t+j} + x_{t+k}}{4k}, \quad i=k+1, k+2, \dots, n-k \quad (2)$$

U slučaju parnih pokretnih proseka vrši se centriranje proseka, tako što se računa aritmetička sredina dva uzastopna pokretna proseka.

Nedostaci metode pokretnih sredina

- Neupotrebljiv je za predviđanje ponašanja vremenske serije u budućnosti, jer se gube podaci sa početka i kraja serije.
- Osetljiva na netipične podatke, što može da se umanjuje korišćenjem metode pokretnih proseka sa težinama.
- Nemogućnost određivanja greške.

Primer tradicionalnog računanja pokretnog proseka i postupak računanja u SPSS-u priložen je u dodacima (Dodatak A, 10.1).

3.1.2.2. Izdvajanje sezonske komponente

Sezonska komponenta predstavlja promene nastale usled pojave faktora koji su takvi da deluju na pojavu, čije stanje ispitujemo, u isto vreme u toku godine i po istom šablonu. Sezonska komponenta se meri sezonskim indeksom, koji se može izračunati različitim metodama. Neke od njih su [15]:

- metoda odnosa prema opštem mesečnom, odnosno kvartalnom proseku,
- korigovani metod odnosa prema opštem mesečnom, odnosno kvartalnom proseku,
- metoda odnosa prema trendu,
- metoda odnosa prema lančanim indeksima,
- metoda X-11,
- metoda odnosa prema pokretnim prosecima,
- itd.

U radu je korišćena metoda odnosa prema pokretnim prosecima, jer je implementirana u SPSS-u u postupku sezonskog desezoniranja, te će jedino ova metoda biti detaljnije opisana.

Metoda odnosa prema pokretnim prosecima

Najčešće korišćena metoda i metoda koja će u ovom radu biti korišćena za izdvajanje sezonske komponente jeste metoda odnosa prema pokretnim prosecima. Ona eliminiše trend, ciklus i slučajnu komponentu iz originalnih podataka kako bi se izdvojila sezonska komponenta.

Prema ovoj metodi sezonski indeks se dobija usrednjavanjem proizvoda (u slučaju multiplikativnog modela dekompozicije vremenske serije, koji će u radu biti korišćen) sezonske i slučajne komponente. Proizvod sezonske i slučajne komponente je jednak količniku originalne serije i serije pokretnih proseka.

Postupak računanja sezonskog indeksa sadrži sledeće korake [8]:

1. Računanje pokretnih proseka,
2. Računanje centriranih pokretnih proseka,
3. Računanje specifičnog sezonskog indeksa³ pomoću formule:

$$I_{ij} = \frac{x_{ij}}{\bar{x}_{ij}} \quad (3)$$

gde je:

x_{ij} – vrednost pojave iz originalne serije

³ Specifični sezonski indeks predstavlja meru sezonske varijacije tokom jedne godine [15].

$\overline{x_{ij}}$ – vrednosti centriranih pokretnih proseka

Ovim računanjem se uklanja uticaj trenda na vremensku seriju.

4. Računanje prosečnog (tipičnog) sezonskog indeksa⁴, čime se eliminiše uticaj ciklične i slučajne komponente. Formula po kojoj se računa tipični sezonski indeks je:

$$I_j = \frac{\sum_{i=1}^n I_{ij}}{n-1}, \quad j = 1, 2, \dots, n. \quad (4)$$

5. Računanje korektivnog faktora, kojim se množi svaki tipični sezonski indeks da bi se dobio izravnati prosečan sezonski indeks. Ovo se radi zbog potencijalnih grešaka u zaokruživanju. Korektivni faktor se računa pomoću formule:

$$\text{Correction factor} = \frac{\text{periodičnost}}{\text{zbir svih tipičnih sezonskih indeksa}} \quad (5)$$

6. Pošto se indeksi obično predstavljaju u procentima, svaka dobijena vrednost se množi sa 100.

U dodacima je predstavljen primer ručnog računanja sezonskog indeksa. (Dodatak B, 10.2)

Desezoniranje podataka

Klasična analiza vremenske serije se uglavnom svodi na postupak otklanjanja sezonske komponente. Ovaj postupak se zove sezonsko izravnavanje ili desezoniranje. Eliminacija sezonske komponente omogućava bolju analizu drugih suštinskih karakteristika vremenske serije – ciklusa, trenda i neočekivanih događaja koji su uticali na prirodno ponašanje vremenske serije (slučajne komponente). Da bi se uklonio efekat sezonske komponente, originalna vremenska serija (koja sadrži trend, ciklus, slučajne i sezonske komponente) se deli tipičnim sezonskim indeksom (TSCI/S) [8]:

$$x'_{ij} = \frac{x_{ij}}{I_j}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m \quad (6)$$

Na ovaj način se dobijaju podaci koji imaju samo trend, ciklus i slučajnu komponentu. U dodacima je predstavljen primer i postupak desezoniranja u SPSS-u (Dodatak C, 10.3).

3.2. Korelacija

Skup statističkih metoda kojima se proučavaju jačina i smer međusobne povezanosti dva atributa predstavlja teoriju korelacije. Osnovni pokazatelji korelacionih veza jesu jednačina regresije i koeficijent

⁴ Prosečni ili tipični sezonski indeks predstavlja meru sezonske varijacije za više uzastopnih godina [15].

korelacije. Za ispitivanje oblika zavisnosti i za njihovo projektovanje u budućnost koristi se regresija, a za ispitivanje jačine i smera veze koristi se koeficijent korelacije. U radu će se koristiti koeficijent korelacije. Koeficijent korelacije pokazuje u kojoj su meri promene vrednosti jednog atributa povezane sa promenama vrednosti drugog atributa. Koeficijent korelacije ispituje povezanost, a ne uzročno posledičnu vezu atributa (ovo znači da se koeficijentom korelacije ne može zaključiti da li jedan atribut utiče na drugi). Postoji više različitih koeficijenata korelacije, a najčešće korišćeni su Pirsonov koeficijent korelacije i Spirmanova korelacija ranga. Pirsonov koeficijent korelacije se koristi nad kontinualnim podacima (intervalnim i razmernim), a Spirmanov nad kategoričkim (ordinalnim i nominalnim podacima) i u slučaju kada kontinualni podaci ne ispunjavaju preduslove za Pirsonov koeficijent korelacije o čemu će kasnije biti više reči.

Slede objašnjenja osnovnih pojmova koji važe i za Pirsonov koeficijent korelacije i za Spirmanovu korelaciju ranga.

Određivanje jačine korelacije

Koeficijent korelacije može imati vrednost između -1 i 1. Jačina korelacije se određuje u zavisnosti od toga koliko je vrednost koeficijenta bliža 1 (ili -1). Što je bliža 1 ili -1 to je korelacija jača; što je bliža 0, to je korelacija slabija. Korelacije se može klasifikovati prema jačini na sledeći način:

od 0 do 0,19	neznatna korelacija
od 0,20 do 0,39	slaba korelacija
od 0,40 do 0,59	umerena korelacija
od 0,60 do 0,79	jaka korelacija
od 0,80 do 1,00	veoma jaka korelacija

Određivanje smera korelacije

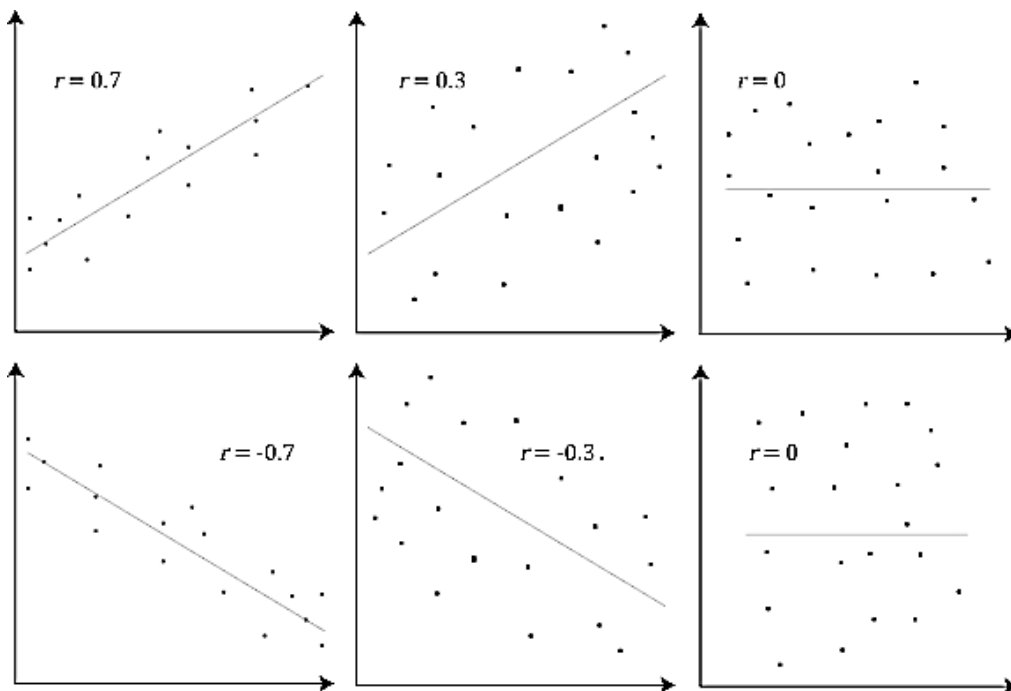
Smer korelacije se određuje na osnovu znaka ispred koeficijenta korelacije. Korelaciju možemo klasifikovati prema smeru u neku od sledećih kategorija [4]:

- **pozitivna korelacija** – kada koeficijent korelacije ima vrednost između 0 i 1. Ovo znači da kada vrednosti jednog atributa raste, vrednost drugog atributa takođe raste.
- **negativna korelacija** - kada koeficijent korelacije ima vrednost između 0 i -1. Ovo znači da kada vrednosti jednog atributa raste, vrednost drugog atributa opada.
- **nema korelacije** - kada koeficijent korelacije ima vrednost 0. Ovo znači da rast vrednosti prvog atributa nema nikakav uticaj na promenu vrednosti drugog atributa.

Ukoliko koeficijent korelacije ima vrednost 1 ili -1, što se u praksi nikad ne dešava, tada se radi o savršenoj korelaciji.

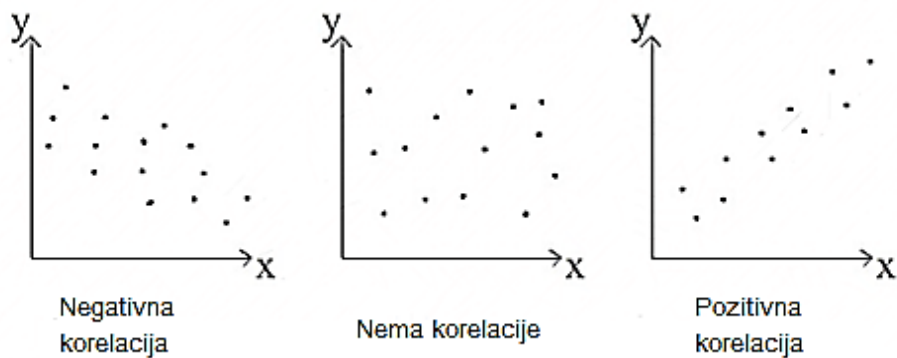
Dijagram rasturanja

Smer i jačina korelacije se mogu utvrditi i na osnovu grafika disperzije, koji se može iscrtati pre primene neke od statističkih metoda za utvrđivanje korelacije. Jačina se može utvrditi na osnovu veličine disperzije (rasturanja) podataka oko regresione linije. Ako atributi nisu povezani, disperzija oko regresione linije je velika. Sa povećanjem linearne povezanosti, disperzija se smanjuje i grafik postaje spljošteniji. Ako između dva atributa postoji apsolutno slaganje onda svi podaci leže na regresionoj liniji. Na slici 6 prikazani su primeri određivanja jačine korelacije preko grafika rasturanja.



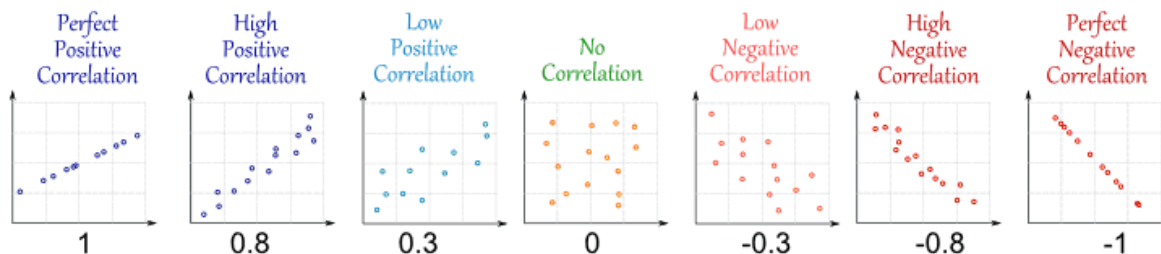
Slika 6. Određivanje jačine korelacije preko dijagrama rasturanja

Na slici 7 su prikazani primeri određivanja smeru korelacije preko dijagrama rasturanja.



Slika 7. Određivanje smeru korelacije preko dijagrama rasturanja

Ako se vrednosti pojave povećavaju sleva na desno korelacija je pozitivna. Ako se vrednosti pojave smanjuju sleva na desno korelacije je negativna. U svim drugim slučajevima korelacija ne postoji. Na slici 8 su prikazani primeri određivanja i smera i jačine korelacije⁵.



Slika 8. Određivanje korelacije preko dijagrama rasturanja

3.2.1. Pirsonov koeficijent korelacije

Pirsonov koeficijent korelacije (r) je koeficijent koji meri jačinu i pravac povezanosti dva atributa koja su izmerena na intervalnoj ili razmernoj skali. Računa se kao količnik kovarijanse i proizvoda standardnih devijacija oba atributa [8]:

$$r_{xy} = \frac{C_{xy}}{SD_x \cdot SD_y} = \frac{\text{kovarijanasa}}{\text{proizvod standardne devijacije atributa } x \text{ i } y} \quad (7)$$

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (7a)$$

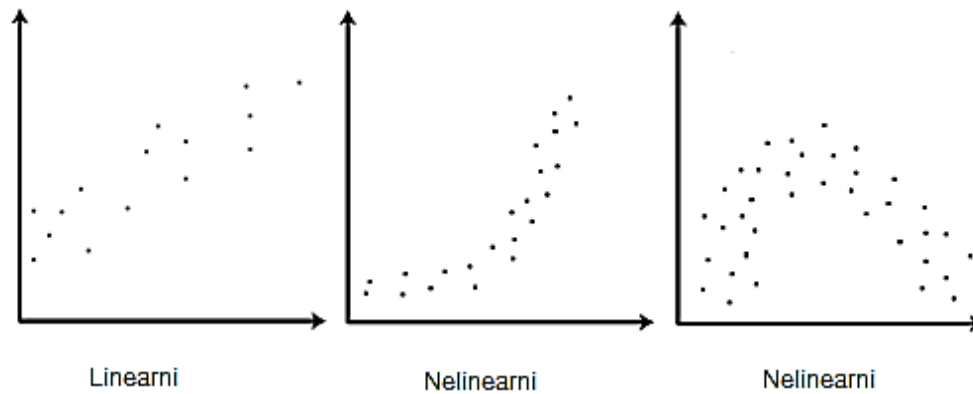
- $(x_i - \bar{x})$ je razlika svake vrednosti atributa x i aritmetičke sredine svih vrednosti atributa x ,
- $(y_i - \bar{y})$ je razlika svake vrednosti atributa y i aritmetičke sredine svih vrednosti atributa y .

Da bi se koristio Pirsonov koeficijent korelacije moraju biti ispunjeni sledeći uslovi:

- Atributi moraju biti izmereni na intervalnoj ili razmernoj skali (kontinualni atributi).
- Vrednosti oba atributa za koja se računa Pirsonov koeficijent korelacije moraju imati normalnu raspodelu. Pre čitanja drugih uslova, poželjno je pročitati dodatak (Dodatak D, 10.4) o normalnoj raspodeli u kojem je dato objašnjenje šta je normalna raspodela i kako se testira upotrebom SPSS-a.

⁵ Izvor: <http://www.mathsisfun.com/data/correlation.html>

- Mora da postoji linearna veza između ta dva atributa, što se može proveriti preko dijagrama rasturanja, fit linije (linije najboljeg poklapanja) ili vrednosti r kvadrata (koeficijenta višestruke determinacije). Na slici 9 dati su primeri linearnih i nelinearnih oblika veze između dva atributa, predstavljeni grafikom rasturanja:



Slika 9. Utvrđivanje postojanja linearne veze preko dijagrama rasturanja

- Takođe, treba imati u vidu i to da je Pirsonov koeficijent korelacije osetljiv na netipične podatke, te bi bilo poželjno očistiti podatke pre računanja ovog koeficijenta.

Testiranje značaja dobijenog koeficijenta

Pošto se koeficijent korelacije izračunava iz uzorka, veoma je bitno pitanje njegovog značaja, tj. pitanja da li je uzorak na osnovu kojeg je koeficijent izračunat dovoljno reprezentativan za donošenje ocene o korelaciji za čitavu populaciju. Sve dok se ne utvrdi stepen značaja, dobijeni koeficijent korelacije predstavlja samo hipotezu. Za utvrđivanje stepena značaja koristi se t-test. Ovim testom se utvrđuje da li je uzorak reprezentativan. Ukoliko jeste, on se može uopštiti i može se doneti zaključak, koji važi za ceo skup podataka.

t-testom se testiraju dve hipoteze: nulta i alternativna [6].

H_0 : korelacija ne postoji; i H_a : korelacija postoji.

Svaki rezultat ovog testa je uvek u vezi sa nultom hipotezom: ili je odbijena ili je prihvaćena. Odbijanjem nulte hipoteze prihvata se alternativna hipoteza i prihvata se izračunata vrednost koeficijenta korelacije iz uzorka kao prava ocena koeficijenta za celokupnu populaciju. Drugim rečima, dobija se rezultat da je uzorak reprezentativan, i da dobijeni rezultat može da se uopšti.

Ako je vrednost koja se dobije t-testom manja od 0.05 (to znači da postoji samo 5% verovatnoće da izmereni koeficijent nije ispravan) nulta hipoteza se odbacuje u korist alternativne hipoteze. Nakon odbijanja nulte hipoteze, može samo da se pretpostavi da je alternativna hipoteza tačna. U slučaju da je

vrednost dobijena t-testom veća od 0.05, nulta hipoteza se prihvata i u tom slučaju alternativna hipoteza ne može biti prihvaćena.

t-test je definisan formulom:

$$t = r_{xy} \frac{\sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \quad (8)$$

gde je:

r_{xy} dobijena koeficijent korelacije iz uzorka,

n je veličina uzorka (broj parova).

Uslovi za primenu t testa:

- Oba atributa koja se testiraju moraju biti kontinualna,
- Ako je veličina uzorka manja od 30 jedinica, vrednosti moraju imati normalnu ili bar simetričnu raspodelu.

3.2.2. Spirmanova korelacija ranga

Spirmanova korelacije ranga (ρ) je neparametarski⁶ ekvivalent Pirsonovog koeficijenta korelacije koji isto kao i Pirsonov koeficijent korelacije meri jačinu veze dva atributa. Razlika je u tome što se računске operacije ne izvode iz numeričkih vrednosti atributa, već iz njihovih relativnih odnosa tj. rangova. Naime, meri se jačina veze dva rangirana atributa. Takođe, razlika je i u tome što Spirmanov koeficijent za razliku od Pirsonovog koeficijenta, ne zahteva da vrednosti atributa budu normalno distribuirani, niti da veza između njih bude linearna.

Postupak određivanja Spirmanovog ranga korelacije se odvija u dve etape [8]:

1. Pre računanja Spirmanovog ranga korelacije, podaci moraju biti rangirani. Stvarne numeričke vrednosti atributa se uređuju po veličini, od najmanje ka najvećoj vrednosti (ili obrnuto) i dodeljuje im se rang od 1 (najvećoj vrednosti u tom setu podataka) do n (broj podataka).
Ovde se ne znaju stvarne razlike između numeričkih vrednosti, već razlike između rangova. Svaki atribut se rangira odvojeno i svakoj njegovoj vrednosti se dodeljuje odgovarajući rang. Ukoliko vrednost atributa za dva podatka ima dve iste vrednosti, onda se uzima aritmetička sredina njihovih rangova (npr. umesto da im se dodeli rang 4 i 5, i jednoj i drugoj se dodeljuje rang 4.5).
2. Kada su podaci transformisani u rangove, pristupa se izračunavanju tzv. ranga korelacije, tj. izračunavanju korelacija među rangovima po formuli:

⁶ Neparametarski - nezavisan od tipa raspodele

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (9)$$

Gde je:

ρ je Spirmanov koeficijent,

d_i je razlika između rangova atributa x i y , za $i = 1, 2, \dots, n$,

n je broj parova vrednosti atributa x i y .

U dodacima je priložen primer računanja Spirmanove korelacije ranga (Dodatak E, 10.5).

Postupak izračunavanja Spirmanove korelacije ranga je kraći i praktičniji od Pirsonove linearne korelacije, pogotovo ako broj parova nije veliki. Međutim, statistička snaga Pirsonovog koeficijenta je znatno veća nego Spirmanove korelacije ranga, pa zato ako su podaci dati intervalno, prednost treba uvek dati Pirsonovom koeficijentu, osim u slučaju kada je narušen neki od uslova za računanje Pirsonovog koeficijenta. I za Spirmanovu korelaciju ranga je neophodno izvršiti test značaja dobijenih rezultata.

Uslovi korišćenja Spirmanove korelacije ranga

Spirmanova korelacija ranga se računa ako je ispunjen neki od sledećih uslova:

- bar jedan od atributa je meren ordinalnom skalom,
- vrednosti ni jednog ni drugog atributa nemaju normalnu raspodelu,
- uzorak je mali,
- veza dva atributa nije linearna.

Napomena: Potrebno je napomenuti i to da visoka korelacija ne znači i postojanje uzročnosti između dva atributa. To što su dva atributa povezana, ne mora nužno da važi i to da jedan atribut uzrokuje drugi. Uvek postoji mogućnost da postoji treći atribut koji uzrokuje menjanje ova dva atributa. Uticaj trećeg atributa može da se proveriti parcijalnom korelacijom [4].

4. Korišćeni softverski alati

U radu su korišćeni softverski alati IBM SPSS Statistics i EViews.

4.1. IBM SPSS Statistics

IBM SPSS Statistics je softverski paket koji se koristi za statističku analizu, obradu i grafički prikaz podataka. Trenutna verzija ovog softvera i verzija koja je korišćena tokom rada jeste 21. Ovaj alat implementira oko 27 različitih algoritama (korelaciju, autokorelaciju, regresiju, ANOVA, MANOVA, ARIMA, naivni Bajesov algoritam, ROC algoritmi...) koji analitičarima pomažu u razumevanju i istraživanju podataka.

U ovom radu IBM SPSS Statistics je korišćen za klasičnu statističku analizu i grafički prikaz vremenskih serija, kao i za analizu korelacije atributa.

4.2. EViews

EViews (Econometric Views) je statistički softver koji ima široku primenu u analizi vremenskih serija. Trenutna verzija je 7.2 i to je verzija koja je korišćena u radu. Za analizu vremenskih serija u EViewsu na raspolaganju su algoritmi: ARIMA, GARCH, Unit root test, test kointegracije, VAR, Multivariant GARCH i Harke-Bera.

U ovom radu **EViews** je korišćen za analizu raspodele podataka pomoću algoritma Harke-Bera.

5. Prikupljanje i skladištenje podataka

Prikupljanje podataka

Podaci koji se koriste u radu i koji su priloženi uz ovaj rad su preuzeti sa navedenih veb stranica:

- dnevni podaci o srednjem deviznom kursu evra i dolara u Republici Srbiji u periodu od januara 2003. – oktobra 2012. godine preuzeti su sa zvaničnog sajta Narodne banke Srbije (<http://www.nbs.rs>)
- Dnevni podaci o kursu evra za dolar u Evropskoj Uniji u periodu od januara 2003. – oktobra 2012. preuzeti su sa sajta Evropske komisije: Eurostat (<http://appsso.eurostat.ec.europa.eu>)
- Mesečni podaci o prosečnim neto zaradama u Republici Srbiji od januara 2003. – oktobra 2012. preuzeti su sa zvaničnog sajta Republičkog zavoda za statistiku (<http://www.rzs.rs>)
- Mesečni podaci o prosečnim penzijama u Republici Srbiji od januara 2003. – oktobra 2012. preuzeti od Pokrajinskog fonda za penzijsko i invalidsko osiguranja u Sremskoj Mitrovici.
- Mesečni podaci o prosečnim zaradama u Republici Hrvatskoj, Republici Makedoniji, Republici Bugarskoj, Republici Rumuniji i Republici Sloveniji su preuzeti sa zvaničnih veb stranica statističkih zavoda tih republika.
- Mesečni podaci o deviznim kursevima u Republici Hrvatskoj, Republici Makedoniji, Republici Bugarskoj, Republici Rumuniji i Republici Sloveniji su preuzeti sa zvaničnih veb stranica Nacionalnih banaka tih republika.
- Godišnji podaci o prosečnim penzijama u Republici Hrvatskoj, Republici Makedoniji, Republici Bugarskoj, Republici Rumuniji i Republici Sloveniji su preuzeti sa zvaničnih veb stranica Republičkih fondova za PIO tih republika.
- Mesečni podaci o broju izvršenih privatizacija u periodu od 2003. do 2009. godine. Pošto podaci nisu dostupni na sajtu Agencije za privatizaciju i ne mogu se dobiti putem lične molbe i zahteva za pristup informacijama od javnog značaja, upućenog Sektoru za pravne poslove Agencije za privatizaciju, napravila sam program koji pomoću programa wget preuzima sve dostupne podatke o privatizovanim preduzećima (sa stranice: <http://www.priv.rs/preduzeca>) i pretražuje ih u potrazi za datumom privatizacije koji zapisuje u Excel datoteku. Ovaj program je napravljen korišćenjem programskog jezika Java. Program i detaljnije objašnjenje programa je priloženo u dodacima (10.7). Ideja je bila da se ovi podaci iskoriste za potvrđivanje hipoteze da priliv deviza od privatizacije utiče na apresijaciju domaće valute. Ipak, ovi podaci nisu korišćeni u radu zbog nepotpunosti podataka. Naime, u vreme pisanja rada na sajtu agencije za privatizaciju nisu bili uneseni podaci o privatizacijama za 2010., 2011. i 2012. godinu.

Skladištenje

Za potrebe skladištenja podataka korišćene su radne sveske (eng. *workbook*) napravljene u programu Microsoft Excel 2010. Korišćene su tri radne sveske: DEVIZNI_KURS, ZARADA i PENZIJA koje se sastoje od kolona predstavljenih na slici 10.

DEVIZNI_KURS	ZARADA	PENZIJA
- Datum	- Datum	- Datum
- EUR_RSD	- neto_zarade	- iznos_penzije
- USD_RSD	- CRO_zarada	
- EUR_USD	- MK_zarada	
- EUR_HRK	- BG_zarada	
- USD_HRK	- RO_zarada	
- EUR_HUF	- HU_zarada	
- USD_HUF	- SLO_zarada	
- EUR_MKD		
- USD_MKD		
- EUR RON		
- USD RON		

Slika 10. Struktura radnih sveski korišćenih za skladištenje podataka

Kao što se sa slike 10 vidi radna sveska DEVIZNI_KURS sadrži podatke o svim deviznim kursovima korišćenim u ovom radu:

- Kolona EUR_RSD sadrži mesečne vrednosti koje je domaća valuta RSD imala za 1 EUR od januara 2003. do oktobra 2012. godine. Ova kolona predstavlja razmerni atribut.
- Kolona USD_RSD sadrži mesečne vrednosti koje je domaća valuta RSD imala za 1 USD od januara 2003. do oktobra 2012. godine. Ova kolona predstavlja razmerni atribut.
- Kolona EUR_USD sadrži mesečne vrednosti koje je 1 USD imao za 1 EUR od januara 2003. do oktobra 2012. godine. Ova kolona predstavlja razmerni atribut.
- Slično je i sa ostalim kolonama: EUR_HRK, USD_HRK, EUR_HUF, USD_HUF, EUR_MKD, USD_MKD, EUR RON, USD RON.

Druga radna sveska – ZARADA, sadrži podatke o zaradama:

- Kolona neto_zarade sadrži prosečne mesečne neto zarade zaposlenih u Republici Srbiji u periodu od januara 2003. do oktobra 2012. godine. Ova kolona predstavlja razmerni atribut.
- Kolona CRO_ZARADE sadrži prosečne mesečne neto zarade zaposlenih u Republici Hrvatskoj u periodu od januara 2003. do oktobra 2012. godine. Ova kolona predstavlja razmerni atribut.
- Slično je i sa ostalim kolonama: MK_ZARADA, BG_ZARADA, RO_ZARADA, HU_ZARADA, SLO_ZARADA.

Treća radna sveska – PENZIJA, sadrži podatke o penzijama:

- kolona `iznos_penzije` sadrži prosečne mesečne penzije u Republici Srbiji u periodu od januara 2003. do oktobra 2012. godine. Ova kolona predstavlja razmerni atribut.

Sve radne sveske imaju i kolonu *datum* koja predstavlja datume merenja posmatranih pojava.

6. Upoznavanje sa podacima

U ovom odeljku predstavljene su uopštene informacije o podacima čija se veza ispituje: deviznom kursu, zaradi i penziji.

6.1. Devizni kurs

Def. Devizni kurs je cena neke strane valute izražena u domaćoj valuti [9].

Domaća valuta može biti u stanju:

- precenjenosti – kada je kupovna moć valute u inostranstvu veća nego u zemlji; i
- potcenjenosti – kada je kupovna moć valute u inostranstvu manja nego u zemlji.

Promene valute

- **Apresijacija** je pojava jačanja vrednosti valute u odnosu na drugu valutu.



Apresijacija domaće valute, odgovara uvoznicima (jer roba iz uvoza postaje jeftinija od domaće), zaposlenima koji primaju platu u dinarima (jer im u konvertovana u evro, plata deluje veće), dužnicima čiji su krediti u stranoj valuti (manje rate), dinarskim štedišama, stranim bankama u Srbiji (mogu jeftino da kupe devize, a onda da povuku novac iz Srbije u svoje matične inostrane filijale), NBS-u (jer će jači dinar možda zaustaviti povlačenje kapitala), Vladi (jer su građani manje nezadovoljni i jači dinar fiktivno povećava BDP⁷ iskazan u evrima i fiktivno se popravља iznos zaduženosti).



Apresijacija valute ne odgovara domaćim proizvođačima koji svoju robu prodaju u inostranstvu, jer postaje skuplja.

- **Depresijacija** je pojava slabljenja vrednosti valute u odnosu na drugu valutu.

⁷ Bruto domaći proizvod



Depresijacija valute utiče na pojeftinjenje domaće robe u inostranstvu, a poskupljenje uvozne robe u zemlji. Slab dinar odgovara dominantnim neto izvoznica, onima koji imaju prihode vezane za evro, bankama koje depozite zasnivaju na pozajmicama iz inostranstva i deviznim štedišama.

Depresijacija ne odgovara uvoznicima i deviznim dužnicima (posebno onima koji su preuzeli kreditne obaveze u stranoj valuti).

▪ **Devalvacija** predstavlja slabljenje vrednosti nacionalne valute nastalo odlukom nacionalne monetarne vlasti. Razlika između depresijacije i devalvacije je u tome što je depresijacija slabljenje vrednosti nacionalne valute nastalo zbog kretanja na deviznom tržištu, a devalvacija odlukom nacionalne monetarne vlasti. Devalvacija je veća promena, dok je depresijacija mala, postepena promena.

Devalvacije se preduzima kada je domaća valuta precenjena, jer to dovodi do deficita u platnom bilansu, većeg uvoza i nedovoljnog izvoza. Devalvacija bi trebalo da olakša otplate državnih inostranih dugova, ali ne u smislu da se teret duga smanji, već da se devalvacijom povećava priliv deviza i smanji njihov odliv, pa će se lakše obezbediti sredstva za otplatu duga.

▪ **Revalvacija** predstavlja jačanje vrednosti nacionalne valute nastalo odlukom nacionalne monetarne vlasti. Razlika između depresijacije i revalvacije je u tome što je depresijacija jačanje vrednosti nacionalne valute nastalo zbog kretanja na deviznom tržištu, a revalvacija jačanje vrednosti nastalo odlukom nacionalne monetarne vlasti. Revalvacija se preduzima u slučaju potcenjenosti domaće valute.

Termini devalvacija i revalvacija upotrebljavaju se u slučaju kada zemlja ima fiksni devizni kurs. A depresijacija i depresijacija se vezuju za sistem fleksibilnih deviznih kurseva.

Vrste deviznih kurseva prema načinu formiranja

a) Fleksibilan devizni kurs

Prema fleksibilnom deviznom kursu vrednost valute se određuje u odnosu na ponudu i potražnju te valute na deviznom tržištu.

Prednosti fleksibilnog kursa:

- daje mogućnost reagovanja na promene situacije nastale na deviznom tržištu,
- manja je mogućnost da dođe do precenjenosti ili potcenjenosti nacionalne valute,
- manji udar na plate, jer se kod fiksnog kursa smanjenje potrošnje mora obaviti preko smanjenja plata i zaposlenih.

Mane fleksibilnog kursa:

- veća je izloženost domaće privrede raznim potresima na svetskom tržištu,
- veći je rizik u poslovanju sa inostranstvom, što utiče negativno na sklapanje dugoročnih poslova sa inostranim partnerima,
- podstiče inflaciju u domaćoj privredi, jer bi zbog znatne depresijacije dinara došlo do drastičnog poskupljenja uvoza i rasta domaćih cena,
- zbog nepredvidivosti, veće su mogućnosti špekulacija.

b) Fiksni devizni kurs

Kod fiksnog deviznog kursa vrednost domaće valute određuje država, tako što je veže za neku drugu referentnu meru vrednosti (npr. zlato, korpu proizvoda, itd.) [9]. U ovom slučaju, država ima obavezu da interveniše kupovinom/prodajom na deviznom tržištu i na taj način „odbrani“ svoj devizni kurs.

Prednosti fiksnog deviznog kursa:

- pozitivno utiče na sklapanje dugoročnih poslova sa inostranim partnerima, jer se u ovom sistemu kompanije ne plaše da će izgubiti na promeni kursa,
- minimizuje špekulacije,
- štiti domaću privredu od iznenadnih promena na svetskom finansijskom tržištu,
- doprinosi stabilizaciji cena u zemlji,
- podstiče uvoz.

Mane fiksnog deviznog kursa:

- zahteva veće devizne rezerve za odbranu kursa,
- ugrožavaju izvoz,
- domaća valuta je često precenjena.

Od zemalja u okruženju fiksiran devizni kurs imaju Hrvatska i BiH.

Faktori koji utiču na promenu deviznih kurseva

Devizni kurs je veoma promenljiv, pa je stalno aktuelno pitanje zašto je u određenom momentu došlo do njegovog pada ili rasta. Neki od glavnih faktora koji utiču na stalne fluktuacije deviznog kursa su: inflacija, javni dug, kretanje kamatne stope, stanje platnog bilansa, rast bruto domaćeg proizvoda, (ne)zaposlenost, stanje deviznih rezervi, i dr.

Devizni kurs u Srbiji

U Srbiji se od 2003. godine primenjuje kontrolisani fleksibilni režim formiranja deviznog kursa. Narodna Banka Srbije jednom dnevno, na kraju svakog radnog dana, slobodno formira i objavljuje svoju kursnu listu za devize i kursnu listu za efektivan strani novac, u zavisnosti od trenutne ponude i potražnje na tržištu novca. Kontrolisani režim znači da Narodna banka Srbije učestvuje na deviznom tržištu, pokušavajući da kupoprodajom deviza utiče na kurs na način koji će onemogućiti drastične promene vrednosti dinara. Nažalost, sve se češće postavlja pitanje da li su devizne rezerve Republike Srbije dovoljno velike da pokriju učestalo subvencionisanje precenjenog dinara i koliko će još dugo biti

sredstava za ovakve poduhvate. Pogotovo ako se u obzir uzme činjenica da se devizne rezerve već godinama ne popunjavaju na osnovu zdravog razvoja domaće privrede, kao što bi trebalo, nego na osnovu novca dobijenog od privatizacija tj. prodaje državne imovine, devizne štednje građana i priliva deviza po osnovu inostranih zaduživanja (kredita) [14].

6.2. Zarada

Prema članu 105. Zakona o radu, zarada predstavlja novčanu dobit koju zaposleni dobija za obavljeni rad i vreme provedeno na radu. Ova dobit sadrži porez i doprinose koji se plaćaju za zaposlenog, uvećanu zaradu, naknadu zarade i druga primanja (osim: naknade troškova za dolazak i odlazak sa rada, za vreme provedeno na službenom putu u zemlji i inostranstvu, otpremnine pri odlasku u penziju, solidarne pomoći, jubilarne nagrade, naknade troškova pogrebnih usluga i naknade štete zbog povrede na radu ili profesionalnog oboljenja) [11].

Razlikuje se neto i bruto zarada. Neto zarada je zarada bez poreza i doprinosa. Bruto zarada je zarada po članu 105. Zakona o radu, koji je prethodno naveden.

Prosečne mesečne zarade se računaju tako što se ukupan iznos isplaćenih zarada u datom mesecu podeli sa brojem zaposlenih na kraju tog meseca.

U analizu koja je izvršena u radu su uključene samo prosečne mesečne neto zarade.

6.3. Penzija

Penzija je mesečna novčana naknada na koju pojedinac stiče pravo na osnovu osiguranja u slučaju starosti, invalidnosti i smrti, a pod uslovima utvrđenim Zakonom o penzijskom i invalidskom osiguranju [13]. Prema članu 18. Zakona o penzijskom i invalidskom osiguranju predviđene su sledeće vrste prava koje pojedinac može ostvariti [13]:

- za slučaj starosti - pravo na starosnu penziju,
- za slučaj invalidnosti - pravo na invalidsku penziju,
- za slučaj smrti:
 - pravo na porodičnu penziju,
 - pravo na naknadu pogrebnih troškova,
- za slučaj telesnog oštećenja prouzrokovanog povredom na radu ili profesionalnom bolešću - pravo na novčanu naknadu za telesno oštećenje,
- za slučaj potrebe za pomoć i negu drugog lica - pravo na novčanu naknadu za pomoć i negu drugog lica.

Zakonom o PIO predviđene su tri kategorije obavezno osiguranih lica:

1. zaposleni,
2. lica koja samostalno obavljaju delatnost, i

3. poljoprivrednici.

Osiguranici zaposleni, osiguranici samostalnih delatnosti i osiguranici poljoprivrednici prava iz penzijskog i invalidskog osiguranja ostvaruju u Republičkom fondu za penzijsko i invalidsko osiguranje i Pokrajinskom fondu kao organizacionoj jedinici tog fonda.

U podatke, korišćene u analizi izvršenoj u radu, su uključene prosečne mesečne starosne penzije za kategoriju korisnika - zaposleni.

7. Priprema podataka

U fazi pripreme izvršeni se sledeći zadaci:

- popunjavanje nedostajućih vrednosti,
- otkrivanje netipičnih podataka,
- transformisanje dnevnih podataka u prosečne mesečne proseke,
- transformisanje atributa čije vrednosti nemaju normalnu raspodelu.

7.1. Popunjavanje nedostajućih vrednosti

U trenutku kada su preuzeti, podaci o deviznim kursevima nisu imali zabeležene podatke o vrednostima kursa tokom vikenda. Kako bi se popunile ove nedostajuće vrednosti napisan je program koji pristupa radnoj svesci i popunjava nedostajuće vrednosti, tako što se svakom danu vikenda dodeli vrednosti deviznog kursa koja je važila petka koji je prethodio tom vikendu. Program je napravljen korišćenjem programskog jezika Java, pri čemu je korišćen *Apache POI Framework* koji se koristi za rad sa Excel datotekama.

U dodacima (Dodatak F, 10.6), na kraju rada, predstavljen je kod sa komentarima.

7.2. Otkrivanje netipičnih podataka

Netipični podaci su podaci koji odstupaju od velike većine podataka. Oni imaju veliki uticaj na rezultat, jer mogu znatno izmeniti rešenje, zbog čega ih je veoma bitno identifikovati i po mogućnosti otkloniti. Popularna su dva pristupa otkrivanja netipičnih podataka pomoću SPSS-a [6]:

- Pravilo identifikovanja netipičnih podataka (Hoaglin i Iglewicz) koje je zasnovano na pravilu identifikovanja netipičnih vrednosti koje je predstavio Tukey 1977. godine. Ovo pravilo se zasniva na množenju interkvartilnog opsega (IQR) faktorom $g=1.5$. **Interkvartilni raspon** (IQR – interquartile range) je razlika između 75-tog i 25-tog kvartila⁸ [4]:

$$IQR = Q_3 - Q_1 \quad (10)$$

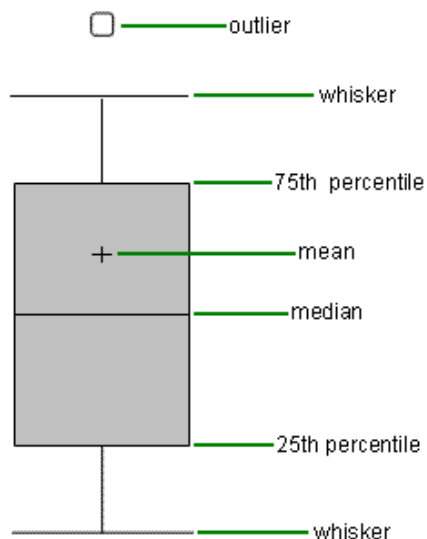
Gornja i donja granica vrednosti koje odvajaju tipične podatke od netipičnih se računa po formuli:

$$(Q_1 - g(Q_3 - Q_1), Q_3 + g(Q_3 - Q_1)) \quad (11)$$

Sve što je van tih granica smatra se netipičnim podatkom.

⁸ Uzorački p-procentni kvantil je vrednost atributa koja je veća ili jednaka od p% elemenata (tj. vrednosti) iz uzorka. [4]

- Ista ova identifikacija netipičnih podataka može da se dobije, bez ručnog računanja, preko vizuelnog prikaza - kućica. Na slici 11 je predstavljeno šta se smatra za netipičnim podatkom, a šta ne.



Slika 11. Otkrivanje netipičnih podataka na grafiku kućice

7.2.1. Otkrivanje netipičnih podataka u atributu EUR_RSD

Samo za otkrivanje netipičnih podataka u atributu EUR_RSD biće korišćeno ručno identifikovanje netipičnih vrednosti kako bi se pokazao postupak, a u svim ostalim slučajevima grafički prikaz - kućice.

Preko opcije Analyze → Descriptive Statistics → Explore, mogu se dobiti informacije o kvartilima, pošto je ova informacija neophodna kako bi se primenilo pravilo identifikovanja netipičnih vrednosti (Tabela 2).

		Percentiles						
		Percentiles						
		5	10	25	50	75	90	95
Weighted		65.0920	68.1220	78.9175	83.7800	99.6950	106.2150	113.4365
Average(Definition 1)	EUR_RSD							
Tukey's Hinges	EUR_RSD			78.9600	83.7800	99.6900		

Tabela 2. Tabela percentila za atribut EUR_RSD

Vrednosti 25. i 75. kvartila su:

$$Q_1 = 78.9600 \text{ i } Q_3 = 99.6900$$

Računaju se donja i gornja granica:

$$Q_1 - g(Q_3 - Q_1) = 78.9600 - 1.5 * 20.7300 = 78.9600 - 31.0950 = 47.8650$$

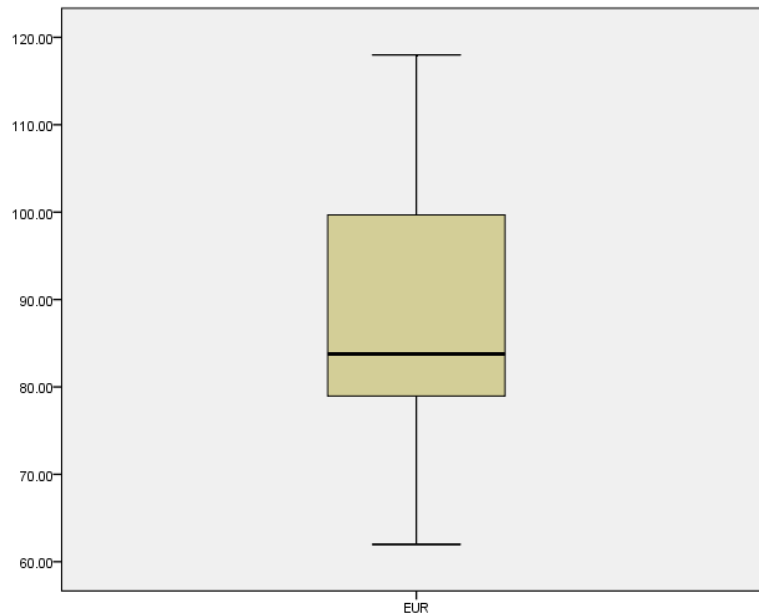
$$Q_3 + g(Q_3 - Q_1) = 99.6900 + 31.0950 = 130.7850$$

Prema tome, donja granica atributa EUR_RSD koja se neće smatrati netipičnim podatkom je 47.8650, a gornja 130.7850. Iz tabele 3 se zaključuje da najveća i najmanja vrednost iz skupa podataka nisu izvan ovih granica, jer je najveća vrednost dinara za jedan evro bila 117.98, a najmanja 64.94. Donosi se zaključak da atribut EUR_RSD nema netipičnih vrednosti.

Extreme Values			
		Case Number	Value
EUR_RSD	1	116	117.98
	2	117	116.51
	Highest 3	115	116.15
	4	114	115.81
	5	118	113.94
	1	1	61.99
	2	2	62.80
	Lowest 3	3	63.61
	4	4	64.15
	5	7	64.94

Tabela 3. Ekstremne vrednosti za atribut EUR_RSD

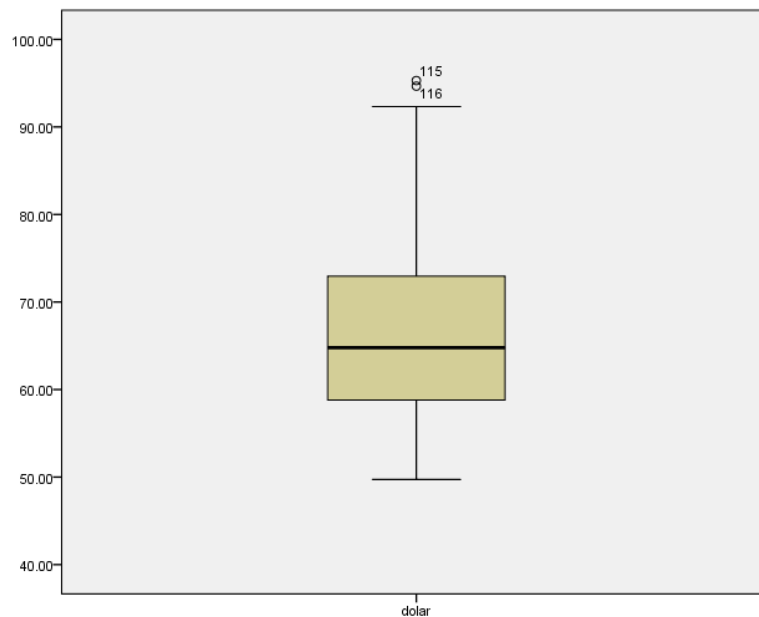
Isti zaključak se donosi i ako se pogleda na grafik kućice (Slika 12).



Slika 12. Grafički prikaz atributa EUR_RSD

7.2.2. Otkrivanje netipičnih podataka u atributu USD_RSD

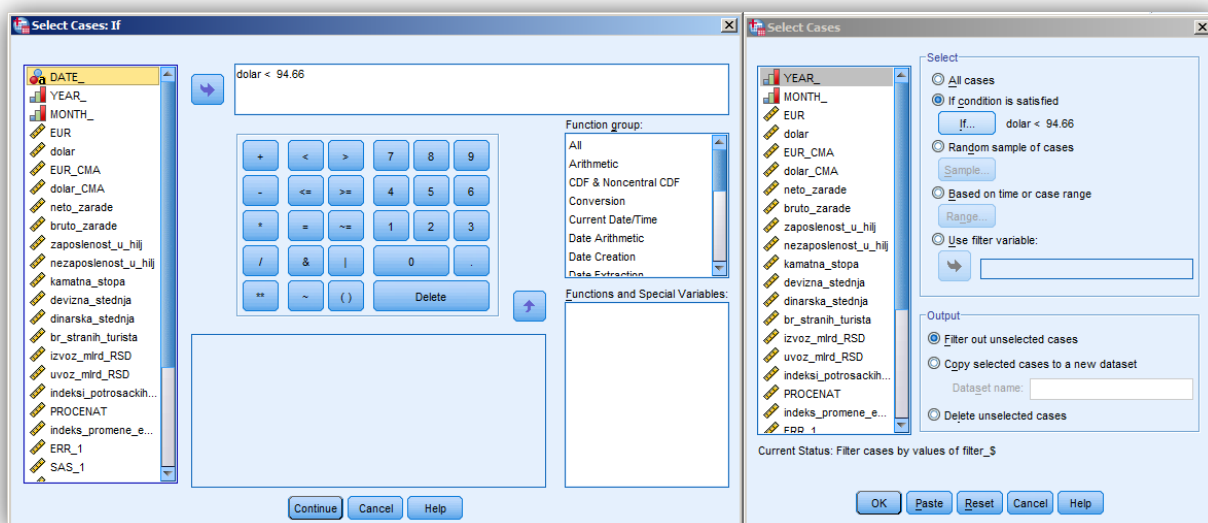
Da bi se utvrdilo da li vrednosti atributa USD_RSD imaju normalnu raspodelu koristi se grafički prikaz kućice (Slika 13).



Slika 13. Grafički prikaz atributa USD_RSD

Na grafiku se vidi da postoje dve netipične vrednosti – u redu 115 i 116. Reč je o julu i avgustu 2012., kada je vrednost dinara za jedan dolar iznosila 94.66 i 95.30 dinara. Kako bi kasniji rezultati ispitivanja

korelacije bili ispravni ove vrednosti moraju biti uklonjene iz analize. Ovo se radi preko opcije Data → Select Cases, gde se postavlja uslov da se uzmu samo vrednosti manje od 94.66 (Slika 14).



Slika 14. Uklanjanje netipičnih podataka iz atributa USD_RSD

7.2.3. Otkrivanje netipičnih vrednosti u atributu neto_zarade

Da bi se utvrdilo da li vrednosti atributa neto_zarade imaju netipične vrednosti koristi se grafički prikaz kućice (Slika 15).

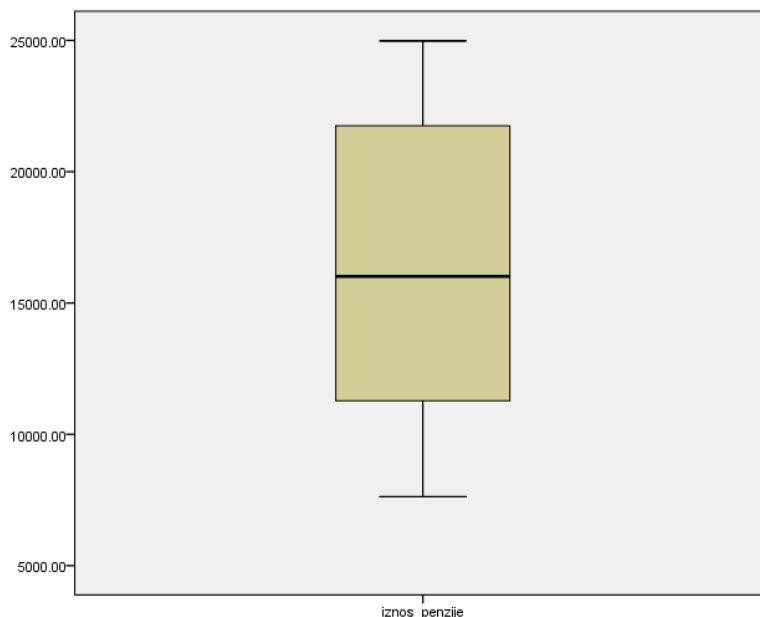


Slika 15. Grafički prikaz atributa neto_zarade

Na osnovu kućica se zaključuje da atribut `neto_zarade` nema netipičnih podataka.

7.2.4. Otkrivanje netipičnih vrednosti u atributu `iznos_penzije`

Da bi se utvrdilo da li vrednosti atributa `iznos_penzije` imaju netipične podatke koristi se grafički prikaz kućice (Slika 16).



Slika 16. Grafički prikaz atributa iznos penzije

Na osnovu grafika kućice se zaključuje da atribut `iznos_penzije` nema netipičnih podataka.

7.3. Transformacija podataka

7.3.1. Transformisanje dnevnih podataka u prosečne mesečne

Pošto je preuzeta baza dnevnih kurseva, a za istraživanje je potrebna baza mesečnih kurseva, jer su intervali penzija i zarada mesečni, mora se izvršiti transformacija dnevnih kurseva u prosečne mesečne.

Transformacija je izvršena u programu Excel preko Pivot tabela.

Postupak se sastoji od navedenih koraka:

1. *Insert -> Pivot Table*,
2. Za *Row Labels* se bira atribut Datum,
3. Za *Values* se bira atribut devizni kurs (EUR_RSD, USD_RSD, EUR_HRK, USD_HRK,...).
4. Za tip kalkulacije se koristi operacija *Average*. Po ovoj operaciji će se vršiti grupisanje.
5. Prosečne vrednosti atributa (deviznog kursa) se grupišu po mesecima. Ovo se radi preko opcije *group->month* nad kolonom datum).

Dobijeni podaci predstavljaju prosečne mesečne vrednosti deviznog kursa.

7.3.2. Transformisanje vrednosti atributa `neto_zarade` u normalno distribuirane podatke

Pošto vrednosti atributa `neto_zarade` nemaju normalnu raspodelu, potrebno je pokušati različitim transformacijama dobiti takve podatke. U te svrhe se primenjuju neke od transformacija:

- $\log$,
- $\sqrt{x}$, i
- $1/x$.

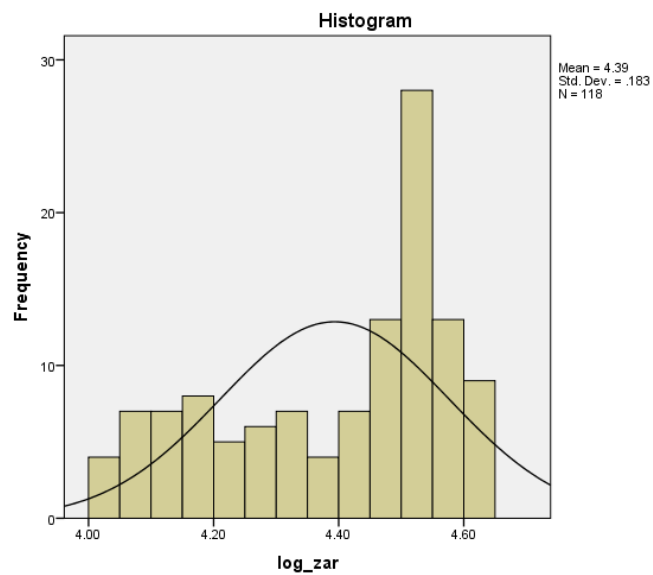
Međutim, proces transformacije vrednosti atributa `neto_zarade` u normalno raspodeljene vrednosti nije bio uspešan. Koeficijenti asimetrije i spljoštenosti su malo poboljšani (u slučaju logaritamske transformacije), ali ako se pogledaju histogrami i rezultat algoritma Jarque-Bera (eng. *Jarque-Bera*)⁹ u EViewsu, videće se da nije došlo ni do kakvih poboljšanja. U tabeli 4 prikazani su koeficijenti asimetrije i spljoštenosti atributa koji predstavljaju transformisane vrednosti atributa `neto_zarade`.

log_zar	Asimetrija	-0.617
	Spljoštenost	-0.937
sqrt_zar	Asimetrija	-0.403
	Spljoštenost	-1.175
rec_zar	Asimetrija	1.079
	Spljoštenost	-0.044

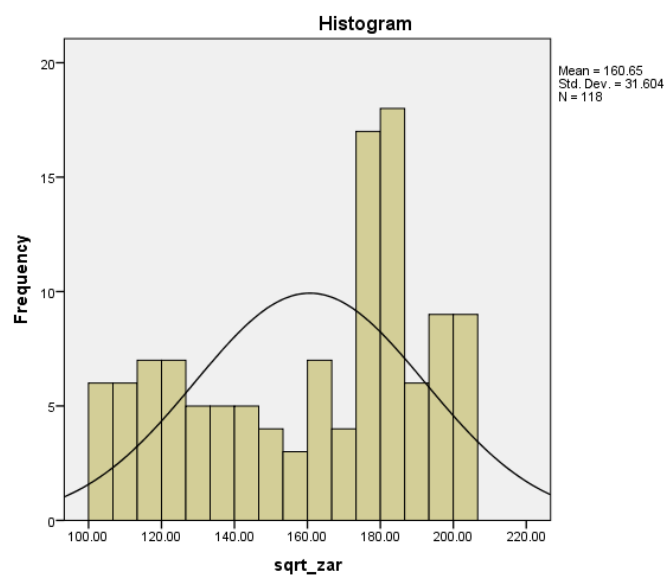
Tabela 4. Koeficijenti asimetrije i spljoštenosti promenljivih koje predstavljaju transformacije atributa `neto_zarade`

Na slikama 17, 18, i 19 su prikazani histogrami ovih promenljivih.

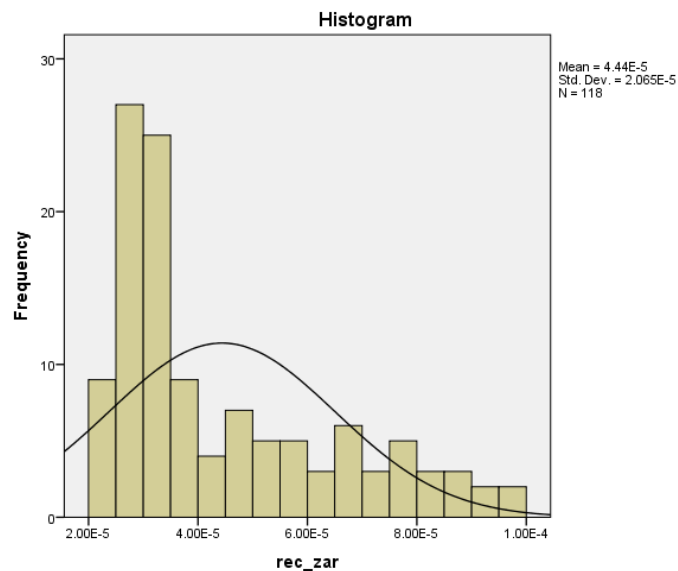
⁹ Jarque-Berin test vrši proveru da li ispitivan uzorak podataka ima vrednosti koeficijenata spljoštenosti i asimetrije koji odgovaraju normalnoj raspodeli.



Slika 17. Histogram atributa `log_zar` koji predstavlja logaritamsku transformaciju atributa `neto_zarade`

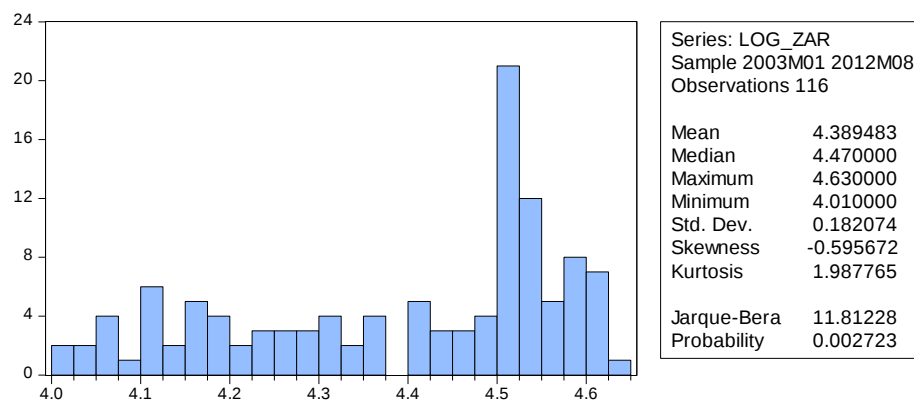


Slika 18. Histogram atributa `sqrt_zar`, koji predstavlja kvadratnu transformaciju atributa `neto_zarade`

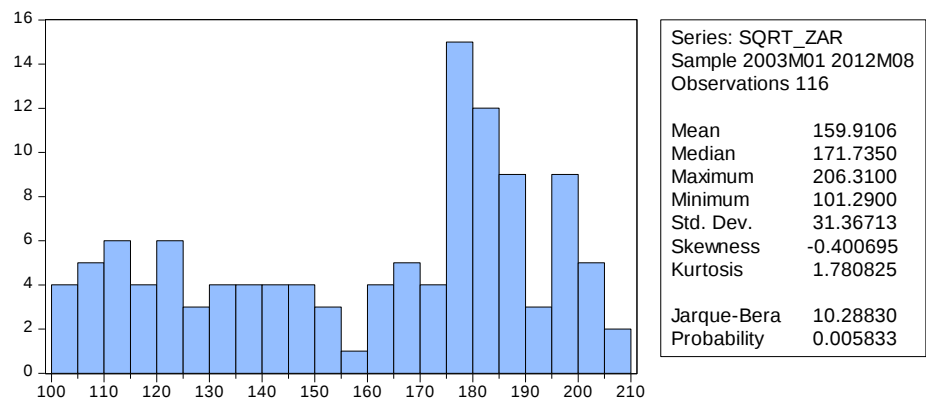


Slika 19. Histogram atributa *rec_zar* koji predstavlja recipročnu vrednost atributa *neto_zarade*

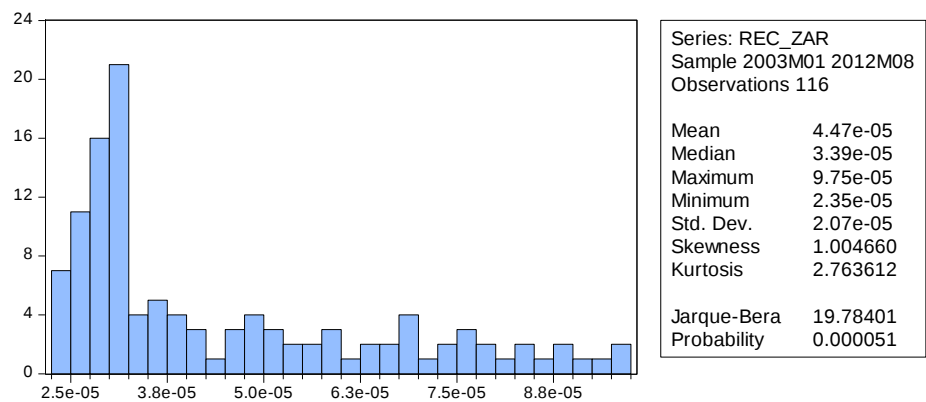
Na slikama 20, 21 i 22 su predstavljeni rezultati primene Harke-Berinog testa na ove attribute.



Slika 20. Rezultat primene Harke-Berinog testa na atribut *log_zar*



Slika 21. Rezultat primene Jarke-Berinog testa na atribut sqrt_zar



Slika 22. Rezultat primene Jarke-Berinog testa na atribut rec_zar

I nakon svih pokušaja transformacija vrednosti atributa neto_zarade i dalje nemaju normalnu raspodelu. Ista je situacija i sa atributima USD_RSD i iznos_penzije, ni nakon primene ovih transformacija nije postignuta normalna raspodela.

8. Karakteristike atributa

U ovom poglavlju biće predstavljena analiza atributa korišćenih u radu. Analiza se sastoji od osnovne deskriptivne statistike, analize raspodele vrednosti atributa, analize trenda i sezone komponente.

8.1. Karakteristike atributa *EUR_RSD*

Za analizu promene vrednosti koju je domaća valuta RSD imala za 1 EUR koristi se atribut *EUR_RSD* iz radne sveske *DEVIZNI_KURS*.

8.1.1. Deskriptivna statistika

U tabeli 5 je predstavljena deskriptivna statistika srednjeg kursa evra. Ona uključuje broj posmatranih vrednosti, najveću i najmanju vrednost koju atribut ima, prosečnu vrednost, koeficijent asimetrije i koeficijent spljoštenosti.

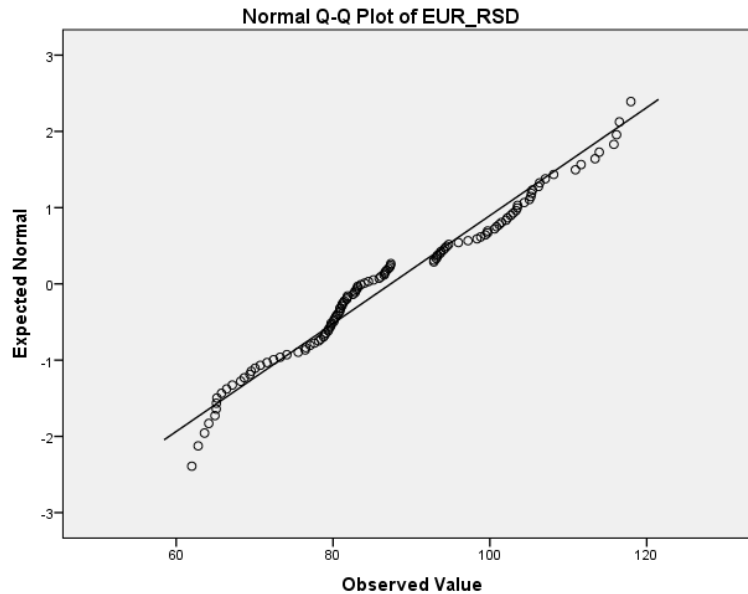
	N	Minimum	Maximum	Mean	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
EUR_RSD	118	61.99	117.98	87.3382	.248	.223	-.761	.442
Valid N (listwise)	118							

Tabela 5. Deskriptivna statistika atributa EUR_RSD

Iz tabele 5 se vidi da je najmanja vrednost dinara za 1 EUR bila 61.99 (u januaru 2003.), a maksimalna 117.98 (u avgustu 2012.). Prosečna vrednost koju je dinar imao za jedan evro je 87.34. Koeficijenti spljoštenosti i asimetrije će biti kasnije analizirani, kada se bude vršila provera raspodele vrednosti atributa.

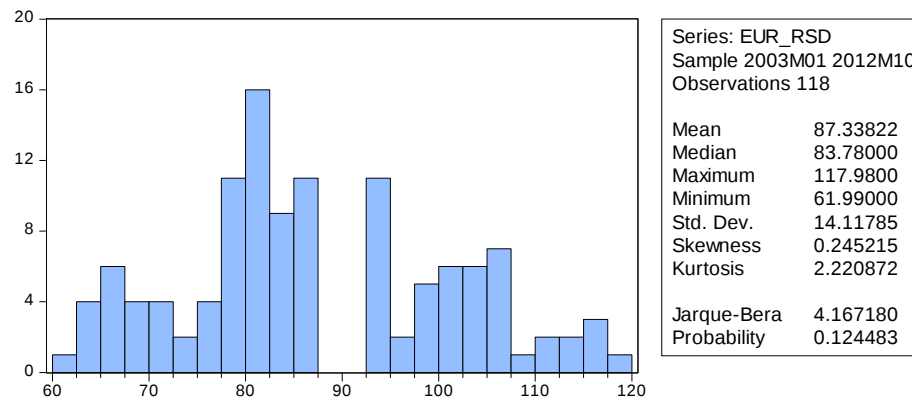
8.1.2. Test normalnosti za atribut *EUR_RSD*

Uzimaju se informacije o koeficijentima asimetrije i spljoštenosti iz tabele 5. Primenom statističkog testa na koeficijent asimetrije dobija se vrednost: $0.248/0.223=1.112$. Ovo znači da se dobijena vrednost koeficijenta asimetrije ne može uzeti u obzir pri testiranju normalnosti atributa. Veoma slična situacija je i kod spljoštenosti. Dakle, raspodela se ne može proveriti na ovaj način. Zbog toga se primenjuje druga provera, preko grafika Normal Q-Q plot (Slika 23).



Slika 23. Normal Q-Q Plot atributa EUR_RSD

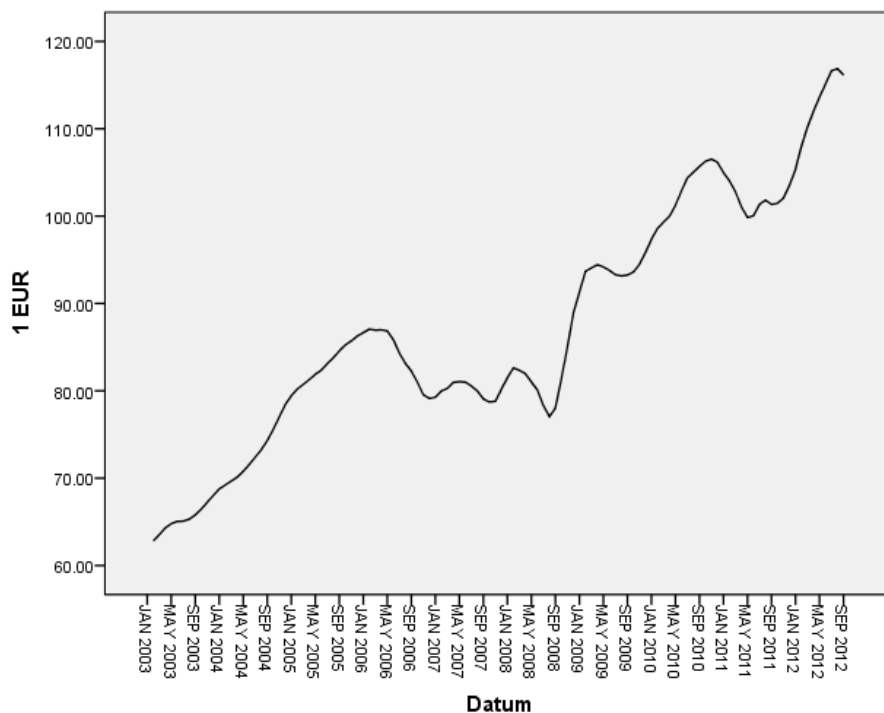
Podaci se većim delom poklapaju sa regresionom linijom, pa bi se moglo zaključiti da atribut EUR_RSD ima normalnu raspodelu. Međutim, pošto je bolje dati prednost nekoj statističkoj metodi za proveru raspodele, nego neobjektivnoj grafičkoj, pogledaćemo i rezultate Harke-Berinog testa u EViewsu. Podsetimo se da su pri proveru raspodele ovim algoritmom da podaci imaju normalnu raspodelu ako je vrednost parametra *Probability* veća od 5%. Prema tome pošto je ovde *Probability* ~12.45%, može se reći da vrednosti atributa EUR_RSD **imaju normalnu raspodelu** (Slika 24).



Slika 24. Prikaz rezultata testa Harke-Bera primenjenog na atribut EUR_RSD

8.1.3. Analiza trenda

Trend promene vrednosti koju je dinar imao za jedan evro u periodu od 2003. do 2012. godine je rastući. To znači da je većim delom ovog perioda vrednost dinara u odnosu na evro depresirala. Na slici 25 prikazan je trend atributa EUR_RSD.



Slika 25. Grafički prikaz trenda atributa EUR_RSD

Sa slike 25 se uočava da je period od januara 2003. do septembra 2006. obeležen neprekidnom depresijacijom dinara - od 2003. do 2006. godine depresirao je za 47%. Tokom ovih godina politička scena Srbije je pretrpela brojne potrese (Vlada je usvojila novi zakon o NBS (14.07.2003.)¹⁰, smena guvernera NBS – Mlađana Dinkića (22.07.2003), neuspeli predsednički izbori (11.2003.), smena guvernerke NBS – Kori Udovički (25.02.2004), formirana nova Vlada Vojislava Koštunice (03.03.2004), usvojena rezolucija o KiM i plan o decentralizaciji KiM (26.03. i 29.04.2004.), izabran novi predsednik – Boris Tadić (27.06.2004.)). Stanje neprekidne depresijacije domaće valute se promenilo u periodu od januara 2006. do februara 2007. godine, kada je dinar ojačao za 8.6%. Uzrok ove apresijacije domaće valute se možda može potražiti u proglašenju nezavisnosti Crne Gore u maju 2006. godine, kao i u održavanju parlamentarnih izbora 21.01.2007. godine. Detaljnijim pregledom vrednosti deviznog kursa tokom ovog perioda, zapaža se da je u avgustu 2006. godine zabeležen pad deviznog kursa evra sa 87.37 RSD za jedan evro, kolika mu je bila vrednost u maju iste godine, na 82.21 RSD. Zanimljivo je da je tog meseca nemačka kompanija Štada izvršila uplatu 475 miliona evra (146.97 EUR po akciji) za akcije akcionarima Hemofarma. Da je ova uplata u dinarima izvršena tri meseca ranije imala bi veću vrednost za 29 miliona evra.

¹⁰ Zakon o Narodnoj Banci Srbije:

http://www.kombeg.org.rs/Slike/UdrFinansijskeOrganizacije/Zakoni/Zakon_o_Narodnoj_banci_Srbije.pdf

Od parlamentarnih izbora do juna 2007. dinar je depresirao za 2.22%, a onda ponovo apresirao pred predsedničke izbore u januaru 2008. Posle izbora (kada je za predsednika izabran Boris Tadić), od januara 2008. do marta 2008. godine, apresirao je za 1.71%. Moguć uzrok ove apresijacije je proglašenje nezavisnosti Kosova i Metohije 17. februara 2008. godine, i održavanju parlamentarnih izbora u maju 2008. godine. U isto vreme dolazi i do svetske krize. Od septembra 2008. do maja 2009. godine, dinar je depresirao za 23.92%. U septembru 2008. godine evro je imao vrednost 76.44 RSD, u decembru 87.32 RSD, a u februaru 2009. čak 93.7 RSD. Zanimljiv podatak je da je u tom periodu, tačnije 24. decembra 2008. godine, potpisan ugovor o prodaji 51% akcija NIS-a ruskoj kompaniji Gaspromnjeft za 400 miliona evra. 02. februara je taj novac i isplaćen. U vreme isplate ugovorena dinarska vrednost akcija je iznosila 372 miliona evra.

Nakon kratke apresijacije, dinar je nastavio da depresira i u periodu od oktobra 2009. do avgusta 2010. godine kada je ekonomska kriza počela snažno da se oseća u Srbiji. U tom periodu dinar je depresirao za 14.92%. Dve godine od početka svetske krize vrednost dinara je oslabila za 39.99%. Ovo je veliki pad vrednosti nacionalne valute u odnosu na evro, pogotovo ako se uporedi sa vrednostima nacionalnih valuta susednih zemalja u tom periodu: Hrvatska kuna je u odnosu na evro u ovom periodu oslabila za samo 0.82%, Makedonski denar za 0.26%, Rumunski lev za 16.58% i Mađarska forinta za 15.33%.

Vratimo se na analizu grafika 25, od novembra 2010. do maja 2011. vrednost dinara je apresirala za 8.13%. Od maja 2011. do avgusta 2012. godine vrednost dinara je depresirala za 19.95%, što je najverovatnije posledica smene Mlađana Dinkića sa funkcije ministra ekonomije, povlačenja političkih kadrova G17+ iz Vlade, izbora nekoliko novih ministara (mart 2011.), političkih izbora održanih 2012. godine i smene guvernera Narodne Banke Srbije (06.08.2012. Jorgovanka Tabaković). U avgustu 2012. dinar dostiže najveću vrednost od 117.98 dinara za jedan evro.

Iz gore-navedenih činjenica može se doneti zaključak da na promenu vrednosti dinara najviše utiče promena unutrašnje i svetske političke klime. Uočava se tradicija da u periodu neposredno pre izbora dolazi do apresijacije dinara, a posle izbora do depresijacije, jer se smatra da je vlast popularnija kad je dinar jači. Uočava se i to da je kroz istoriju često dolazilo do drastičnih promena u kursu neposredno pre i posle privatizovanja značajnijih preduzeća.

8.1.4. Analiza sezonske komponente

Analizom sezonske komponente napravljen je pokušaj utvrđivanja da li postoji određeni period godine kada je apresijacija ili depresijacija deviznog kursa veća u odnosu na druge periode godine (npr. da li leti dinar depresira zbog dolaska turista). U tabeli 6 prikazani su mesečni sezonski indeksi za atribut EUR_RSD.

Seasonal Factors

Series Name: EUR_RSD

Period	Seasonal Factor (%)
1	100.5
2	100.9
3	101.1
4	100.3
5	100.9
6	100.2
7	99.5
8	99.4
9	99.2
10	98.4
11	99.6
12	99.8

Tabela 6. Sezonski indeksi za atribut EUR_RSD

Na osnovu dobijenih rezultata primećuje se da dinar najviše apresira u oktobru, a u martu najviše depresira. Za vreme sezone turista (od juna do oktobra) dinar ima običaj da jača. Pogotovo, u septembru i oktobru, kada je broj turista 30% iznad proseka¹¹, dinar je najjači. Računica je jasna - ako je jedan evro u „sezoni turista“ 100 dinara, turista će moći da proda 100 evra za 10000 dinara, nakon „sezone turista“ kada je evro 103 dinara (od decembra do aprila kada je broj turista znatno manji), neko ko je kupio tih 100 evra može zaraditi 300 dinara. Pošto nijedan mesec nema izražen sezonski faktor nije potrebno raditi desezoniranje.

8.2. Karakteristike atributa USD_RSD

Za analizu promene vrednosti koju je domaća valuta RSD imala za 1 USD koristi se atribut USD_RSD iz radne sveske DEVIZNI_KURS.

8.2.1. Deskriptivna statistika

U tabeli 7 je predstavljena deskriptivna statistika srednjeg deviznog kursa dolara.

	N	Minimum	Maximum	Mean	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
USD_RSD	118	49.72	95.30	66.6171	.727	.223	.014	.442
Valid N (listwise)	118							

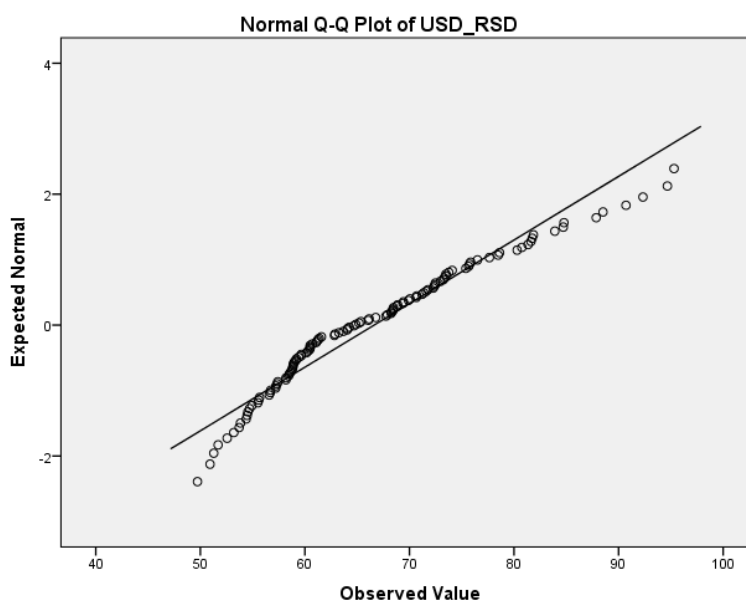
Tabela 7. Deskriptivna statistika atributa USD_RSD

¹¹ Zaključak izveden na osnovu podataka dostupnih na sajtu Republičkog zavoda za statistiku.

Iz tabele 7 se vidi da je najmanja vrednost dinara za jedan dolar bila 49.72 RSD (u julu 2008.), a maksimalna 95.30 RSD (u avgustu 2012., ista situacija kao i kod evra). Prosečna vrednost koju je dinar imao za jedan dolar, u periodu od 2003. do 2012. godine, je 66.62.

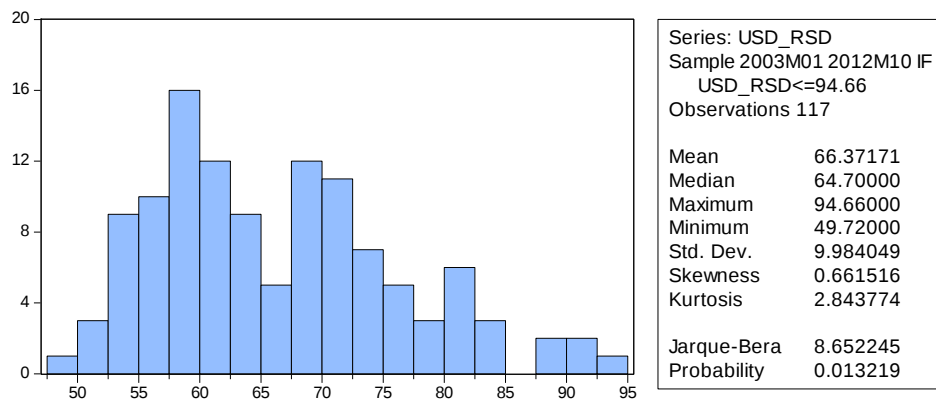
8.2.2. Test normalnosti za atribut USD_RSD

Uzimaju se informacije o koeficijentima asimetrije i spljoštenosti iz tabele 7. Primenom statističkog testa na koeficijent asimetrije dobija se vrednost: $0.727/0.223 = 3.26$, što znači da se dobijena vrednost asimetrije može uzeti u obzir pri testiranju normalnosti vrednosti atributa. Međutim, pošto se vrednost koeficijenta spljoštenosti ne može uzeti u obzir, zbog $0.14/0.442=0.32$, odbacujemo ovaj vid testiranja. Radi se druga proveru, proveru preko grafika Normal Q-Q plot. Na grafiku Normal Q-Q Plot (Slika 26) se vidi da se vrednosti atributa USD_RSD ne poklapaju sa regresionom linijom, te bi na osnovu ovoga moglo da se pretpostavi da atribut USD_RSD nema normalnu raspodelu.



Slika 26. Normal Q-Q Plot za atribut USD_RSD

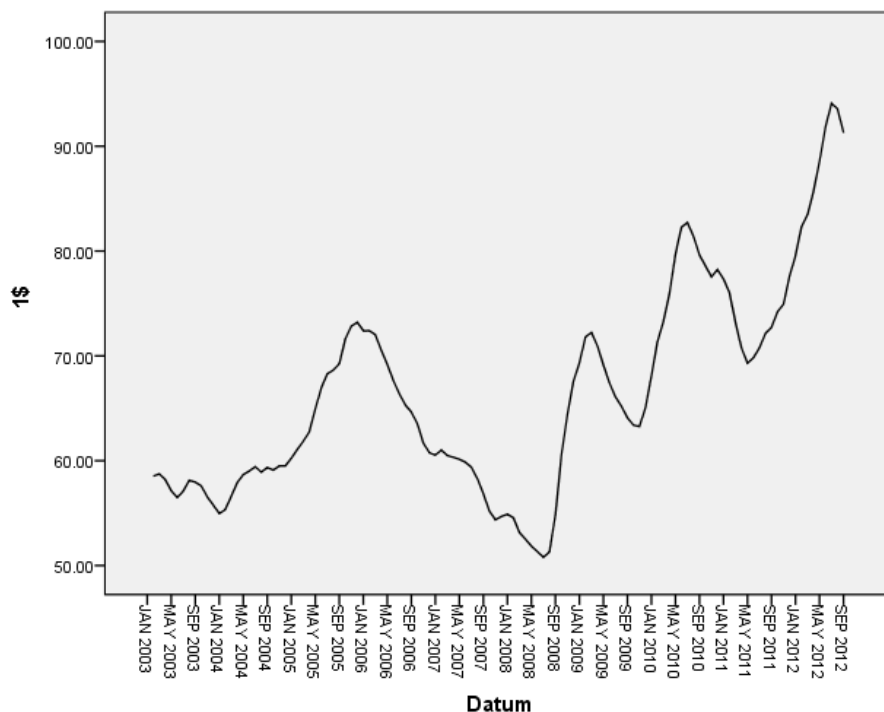
Isti zaključak se izvodi i ako se raspodela proverava preko histograma. Ali, da bi se to potvrdilo, bilo bi dobro pogledati i Harke-Berin test kako bi se donela konačna odluka. Na slici 27 se vidi da je koeficijent *Probability* manji od 5%, te se na osnovu toga donosi zaključak da vrednosti atributa USD_RSD **nemaju normalnu raspodelu**.



Slika 27. Rezultati testa Jarque-Bera za atribut USD_RSD

8.2.3. Analiza trenda

Trend promene vrednosti koju je dinar imao za jedan dolar u periodu od 2003. do 2012. godine nije izražen. Na slici 28 prikazan je trend atributa USD_RSD.



Slika 28. Grafički prikaz trenda atributa USD_RSD

Sa slike 28 se vidi da je vrednost dinara, od januara 2004. do februara 2006. godine, depresirala za 27.02% u odnosu na dolar (podsetimo se da je u isto vreme i evro depresirao), a potom apresirala za 12.38% do septembra 2008 (u ovom periodu je Crna Gora proglasila nezavisnost, održani su

predsednički izbori i KiM je proglasila nezavisnost). Nakon ovog perioda, od septembra 2008. do februara 2009. godine, usledila je znatna depresijacija domaće valute za 43.82%. Tada je dinar depresirao i u odnosu na evro za 23.92%. Pretpostavlja se da je ova depresijacija nacionalne valute prouzrokovana održanim predsedničkim izborima i svetskom krizom. Od februara 2009. do oktobra 2009. vrednost dinara je apresirala u odnosu na dolar za 14.17% (Neka od dešavanja koja su možda uticala na ovu su: započeta primena Prelaznog trgovinskog sporazuma između Srbije i EU u januaru 2009; Potpisan je ugovor o bezviznom režimu između Srbije i Rusije u februaru, a u decembru između Srbije i EU), a onda je opet do juna 2010. depresirala za čak 34.72%.

Vrednost dinara je posle dve godine od početka svetske krize u odnosu na dolar oslabila za 41.38% (za 2% više nego što je dinar oslabio u odnosu na evro). Ovo je veliki pad vrednosti nacionalne valute u odnosu na dolar, što se može zaključiti i ako se uporedi sa vrednostima deviznih kurseva susednih zemalja u tom periodu: Hrvatska kuna je u odnosu na dolar u ovom periodu oslabila za samo 12.02% (za 10% više nego što je oslabila u odnosu na evro), Makedonski denar za 11.66% (za 11% više nego što je oslabio u odnosu na evro), Rumunski lev za 29.08% (za 10% više nego što je oslabio u odnosu na evro) i Mađarska forinta za 20.42% (za 5% više nego što je oslabio u odnosu na evro). Imajući u vidu sve navedeno, donosi se zaključak da je srpski dinar najviše oslabio. Po jačini depresijacije prati ga Rumunski lev, dok su druge valute imale približno isti procenat depresijacije.

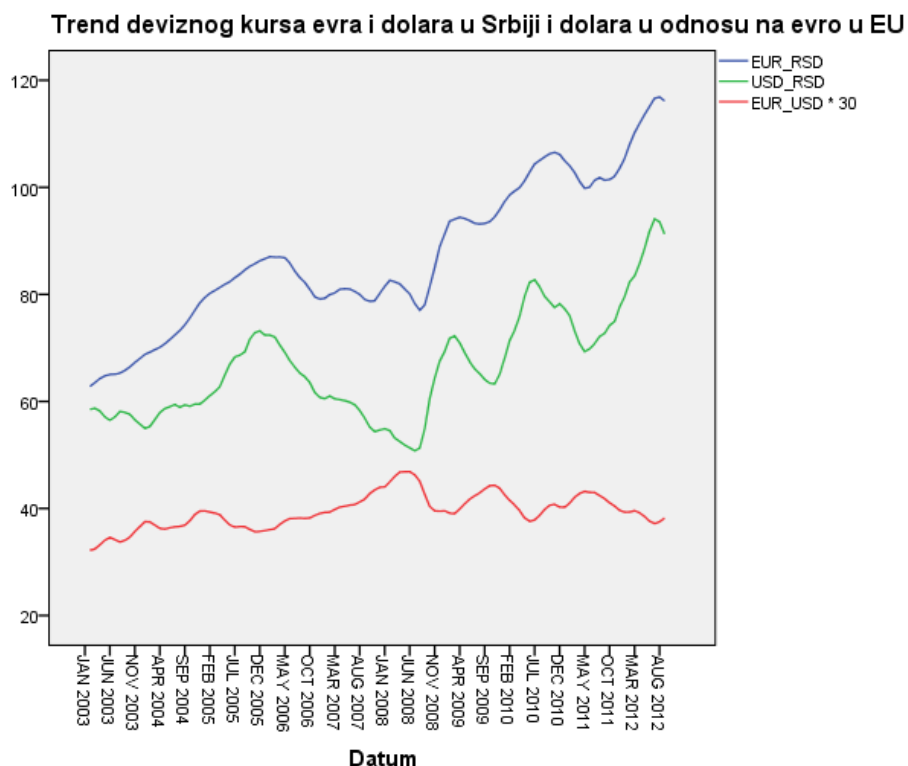
Vratimo se na analizu grafika 28, od juna 2010. do maja 2011. dinar je apresirao u odnosu na dolar za 19.15%, a od maja 2011. do avgusta 2012. godine je depresirao za čak 39.16% (smena ministara u Vladi, predsednički izbori, smena guvernera NBS). U avgustu 2012. je ostvarena najmanja vrednost dinara za jedan dolar - 95.30 dinara. Za samo godinu dana, od avgusta 2011. do avgusta 2012. dinar je depresirao u odnosu na dolar za čak 33.53% (dok je u odnosu na evro depresirao 15.4%).

Trend evra i dolara u EU

Na grafiku 29 je prikazana promena trenda deviznog kursa evra i deviznog kursa dolara u Srbiji, kao i deviznog kursa dolara u EU. Sa ovog grafika može se primetiti da je tokom celog perioda od januara 2003. do septembra 2012. dolazilo do približno istovremenih fluktuacije vrednosti dinara, u odnosu na dolar u Srbiji i dolar u EU. Sa grafika se vidi i to da su USD_RSD i EUR_USD negativno korelisani, što znači da kada dolar jača u EU jača i u Srbiji¹².

¹² Kada dolar jača, na grafiku EUR_USD vrednost opada, i obrnuto kada dolar slabi, na grafiku EUR_USD vrednost raste. Kada dolar jača, na grafiku USD_RSD vrednost raste, i obrnuto kada dolar jača, na grafiku USD_RSD vrednost opada. Kada evro jača, na grafiku EUR_USD vrednost raste, i obrnuto kada evro slabi, na grafiku EUR_USD vrednost opada. Kada evro jača, na grafiku EUR_RSD vrednost raste, i obrnuto kada evro slabi, na grafiku EUR_USD vrednost opada.

Međutim, sa evrom je situacija drugačija. U periodu „zlatnog doba privatizacije“¹³ u Srbiji (od 2003. do 2007.), kada je prodato oko 1400 preduzeća, kurs evra ne prati kurs evra u EU. Kurs evra u Srbiji uglavnom u tom periodu raste tj. dinar slabi. Tek od početka 2007. kurs evra počinje da prati vrednost evra u EU. Međutim ni to ne traje dugo, takvo stanje se zadržava do 2010. godine¹⁴.



Slika 29. Prikaz trenda deviznog kursa evra i dolara

Srbija i zemlje u susedstvu

Kako bi se utvrdilo u kakvoj je poziciji domaća valuta Srbije u odnosu na domaće valute susednih zemalja, data je tabela 8 u kojoj su prikazane prosečne godišnje vrednosti domaće valute, Republike Srbije i zemalja u regionu, za jedan evro/dolar u periodu od 2008. – 2012. godine.

¹³ Najviše preduzeća je privatizovano u 2003. godini, skoro oko 680 preduzeća. Ali, i u decembru 2002. kada je za samo mesec dana prodato 112 preduzeća. Zbog ovog velikog priliva deviza iz 2002. trebalo bi da je dinar u 2003. godini jači, međutim to ipak nije bio slučaj. U ovom periodu guverner Narodne Banke Srbije je bio Mlađan Dinkić (od 11.2000. do 07.2003. godine), a direktor Agencije za privatizaciju Mirko Cvetković (od 2003. do 2004. godine).

¹⁴ U ovom periodu guverner Narodne Banke Srbije je bio Radovan Jelašić (od 02.2004. do 07.2010. godine). A od 2010. do 2012. godine guverner je Dejan Šoškić.

Zemlja	2008	2009	2010	2011	2012
Srbija					
EUR	81.42	93.87	102.85	101.94	113.00
USD	55.70	67.47	77.71	73.32	88.00
Hrvatska					
EUR	7.22	7.34	7.29	7.43	7.52
USD	4.93	5.28	5.50	5.34	5.85
Mađarska					
EUR	251.66	280.33	275.29	279.35	292.03
USD	172.34	202.15	207.98	201.08	225.02
Makedonija					
EUR	61.27	61.27	61.51	61.53	61.57
USD	41.86	44.08	46.46	44.23	47.92
Rumunija					
EUR	3.68	4.23	4.21	4.24	4.45
USD	2.52	3.05	3.18	3.05	3.47
Bugarska					
EUR	1.96	1.96	1.96	1.96	1.96
USD	1.34	1.41	1.48	1.41	1.52

Tabela 8. Prikaz fluktuacije domaće valute Srbije i susednih zemalja u odnosu na evro i dolar

Podaci predstavljeni u tabeli 8 će biti od značaja pri konverziji zarada i penzija u evre i dolare, u poglavlju 8.3.3. Na osnovu ovih podataka može se zaključiti o jačini domaće valute. Može se uporediti procenat depresijacije valuta, što je i urađeno i predstavljeno u tabeli 9. U tabeli 9 je dat procenat depresijacije dinara, hrvatske kune, mađarske forinte, makedonskog denara, rumunskog leja i bugarskog leva u odnosu na evro i dolar.

Zemlja	2008. – 2010.	2010. – 2012.	2008. – 2012.
Srbija			
EUR_RSD	29.92%	19.85%	42.54%
USD_RSD	39.87%	33.20%	63.41%
Hrvatska			
EUR_HRK	0.97%	3.16%	4.16%
USD_HRK	11.96%	6.36%	18.66%
Mađarska			
EUR_HUF	9.39%	6.08%	16.04%
USD_HUF	20.68%	8.19%	30.57%
Makedonija			
EUR_MKD	0.39%	0.09%	0.49%
USD_MKD	10.99%	3.14%	14.48%
Rumunija			
EUR RON	14.40%	5.70%	20.92%
USD RON	26.19	9.12%	37.70%

Bugarska			
EUR_BGN	0%	0%	0%
USD_BGN	10.45%	2.70%	13.43%

Tabela 9. Depresijacija dinara u odnosu na evro i dolar predstavljen u procentima

Iz tabele 9 se zaključuje da je ubedljivo najveću depresijaciju domaće valute u odnosu na evro i dolar imala Republika Srbija. Glavni efekat ove depresijacije domaće valute je smanjenje vrednosti zarada i penzija, izraženih u evrima i dolarima, što će biti pokazano u odeljku 8.3.3.

8.2.4. Analiza sezonske komponente

Radi ispravnije analize odnosa među atributima, uvek je dobro prethodno ispitati sezonski uticaj na vrednost tog atributa i ukoliko se utvrdi da je značajan, ukloniti ga tj. izvršiti postupak desezoniranja.

Seasonal Factors

Series Name: USD_RSD

Period	Seasonal Factor (%)
1	100.6
2	100.9
3	100.6
4	99.7
5	100.8
6	100.4
7	100.0
8	100.0
9	99.3
10	99.1
11	100.1
12	98.5

Tabela 10. Sezonski indeks za atribut USD_RSD

Kao što se vidi iz tabele 10 atribut dinar je najviše apresira u odnosu na dolar u decembru, ali pošto nijedan od ovih sezonskih faktora nije izražen, nad atributom USD_RSD nema potrebe raditi desezoniranje.

8.3. Karakteristike atributa neto_zarade

Za analizu promene vrednosti neto zarada u Republici Srbiji koristi se atribut neto_zarade iz radne sveske ZARADA.

8.3.1. Deskriptivna statistika

U tabeli 11 je predstavljena deskriptivna statistika atributa *neto_zarade*.

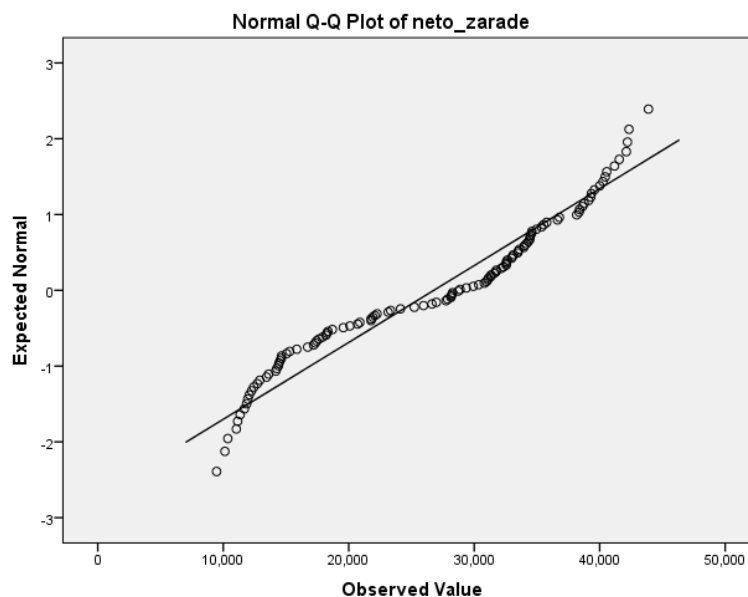
	N	Minimum	Maximum	Mean	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
neto_zarade	118	9468.00	43887.00	26781.1864	-.195	.223	-1.276	.442
Valid N (listwise)	118							

Tabela 11. Deskriptivna statistika atributa *neto_zarade*

Iz tabele 11 se vidi da je najmanja prosečna zarada bila 9468 RSD (u januaru 2003.), a najveća 43887 (u decembru 2011.). Prosečna vrednost prosečne neto zarade u Srbiji u periodu od 2003. do 2012. godine je 26781.19.

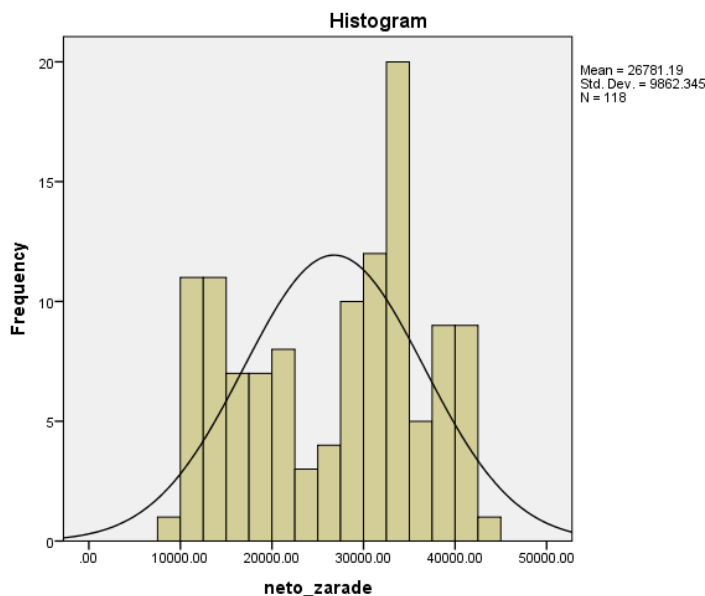
8.3.2. Test normalnosti za atribut *neto_zarade*

Uzimaju se informacije o koeficijentima asimetrije i spljoštenosti iz tabele 11. Primenom statističkog testa na koeficijent asimetrije dobija se vrednost: $0.259/0.223=1.112$, što znači da se dobijena vrednost koeficijenta asimetrije ne može uzeti u obzir pri testiranju normalnosti atributa. Zato se radi druga provera preko grafika Normal Q-Q plot. Pošto se vrednosti atributa *neto_zarade* ne poklapaju sa regresionom linijom zaključuje se da vrednosti atributa *neto_zarade* nemaju normalnu raspodelu (Slika 30).



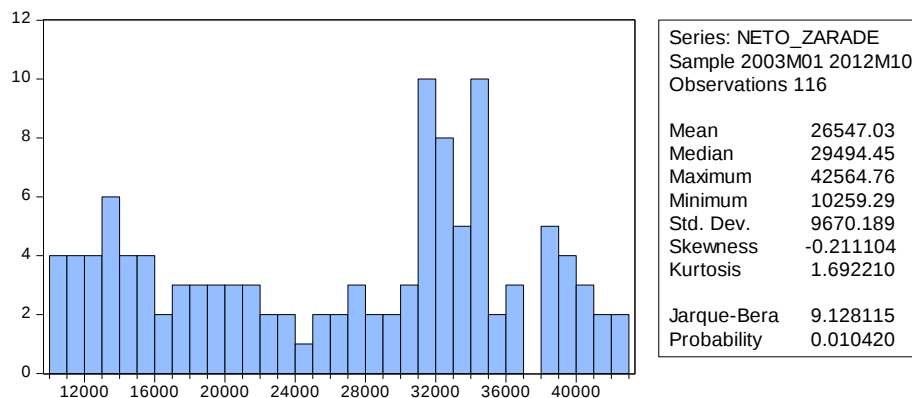
Slika 30. Normal Q-Q Plot za atribut *neto_zarade*

Krivulja normalnosti na histogramu govori suprotno (Slika 31), zbog čega se za donošenje konačne odluke uzima u obzir rezultat testa Harke-Bera.



Slika 31. Histogram za atribut neto_zarade

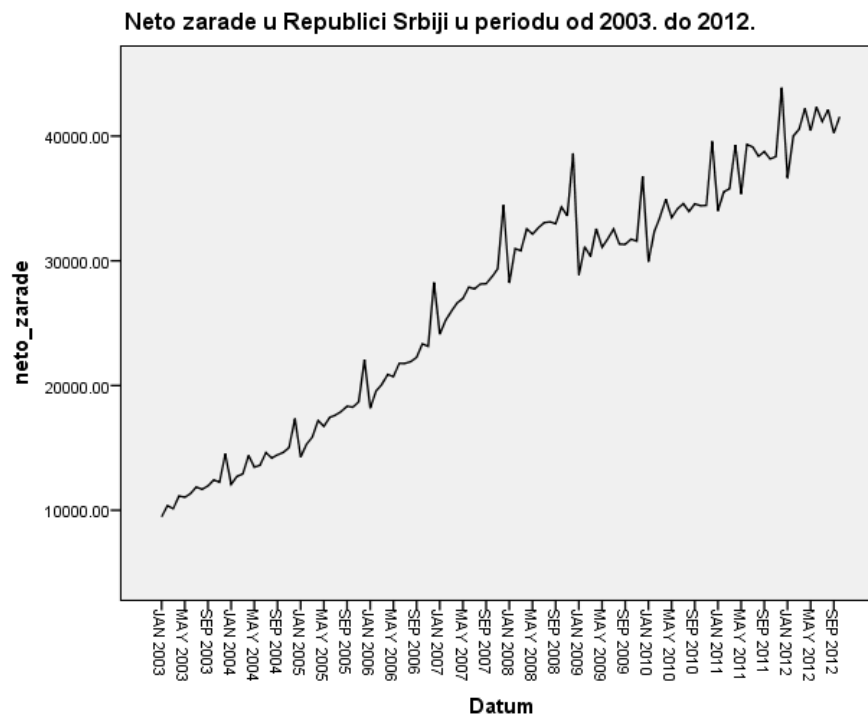
Pošto je vrednost parametra *Probability* manja od 5%, donosi se konačan zaključak da vrednosti atributa neto_zarade **nemaju normalnu raspodelu** (Slika 32).



Slika 32. Rezultat testa Harke-Bera za atribut neto_zarade

8.3.3. Analiza trenda

Sa slike 33 se zaključuje da prosečna mesečna neto zarada u Republici Srbiji, u periodu od 2003. do 2012. godine, ima rastući trend. Takođe, primećuje se da postoji izražen sezonski indeks za mesec decembar, jer je svake godine u decembru zabeležen znatan porast zarada. Ovaj znatan rast vrednosti zarada se javlja zbog trinaeste plate koju zaposleni dobijaju na kraju svake godine.



Slika 33. Trend neto zarade u RS

Sa slike 33 se primećuje i to da je u januaru 2009. godine¹⁵ došlo do pada vrednosti zarade za 33.76%. Tada je zarada bila za 43.92% manja nego tri godine kasnije, u oktobru 2012. Pretpostavlja se da je ovaj pad zarade tj. zamrzavanja zarada i penzija u tom periodu, posledica ekonomske krize, koja je ostavila posledice i na domaću valutu. Domaća valuta je tada pretrpela najveću depresijaciju u odnosu na evro (za ~20%) i dolar (za ~40%), što se može videti u poglavlju 8.2.3.

Da bi se izveo zaključak o tome koliko se brzo menjala plata zaposlenih u Srbiji, izvršeno je poređenje sa vrednostima plata zaposlenih u susednim državama. U tabeli 14 su dati procenti rasta vrednosti zarade u Srbiji i u zemljama u susedstvu za period od 2008. – 2012. godine.

Zemlja	2008. – 2010.	2010. – 2012.	2008. – 2012.
Srbija	4.28%	38.86%	26.34%
Hrvatska	3.19%	2.52%	5.82%
Mađarska	12.12%	12.70%	26.36%
Makedonija	27.70%	1.69%	29.87%
Rumunija	9.70%	9.13%	19.71%
Bugarska	24.23%	18.10%	46.72%
Slovenija	17.07%	2.57%	9.82%

¹⁵ Od 2008. do 14.03.2011. ministarka finansija je bila Diana Dragutinović (DS)

Tabela 12. Procenti promene zarada u susednim zemljama

Prema tabeli 12 se može zaključiti da je Republika Srbija od 2010. do 2012. godine, u odnosu na druge države, imala najveći porast plata. Takođe, izrazit rast je imala i tokom perioda od 2008. do 2012. godine. Međutim, kada se pogleda koliko je depresirala domaća valuta (Tabela 9) i kada se pogleda plata izražena u evrima i dolarima (Tabela 13) donosi se drugačiji zaključak.

Ako se ponovo pogleda tabela 9 uočava se da je dinar od 2008. – 2010. u odnosu na evro (dolar) depresirao za 29.92% (39.87%), dok se iz tabele 12 vidi da je zarada u tom periodu porasla za samo 4.28%. Ista je situacija i za period od 2008. – 2012. godine, kada je dinar depresirao za 42.54% (63.41%), a zarada porasla za samo 26.34%. To znači da rast plata ne prati rast deviznog kursa, što dalje znači da realnog rasta zarade nije ni bilo. Ovo se najbrže uočava ako se pogledaju zarade iz 2008. i 2010. godine. 2008. godine zarada je u dinarima iznosila 32757 dinara, što je za 1402 dinara manje nego 2010. godine. Međutim, izražena u evrima, zarada iz 2008. je za 70 evra bila veća u odnosu na onu iz 2010. godine. Slična situacija je, pa čak i nepovoljnija, sa dolarom - dinar je depresirao u odnosu na dolar za 63.41%, a zarada izražena u dolarima se od 2008. do 2010. smanjila za 152 dolara. Isto tako, kada se izračuna procenat povećanja zarade predstavljene u evrima, dolazi se do drugačijeg zaključka od onog koji je izveden na osnovu tabele 12. Od 2008. do 2010. vrednost plate je smanjena za 6.85%, od 2008. do 2012 plata je smanjena za 9.04%, a od 2010. do 2012. je povećana za 10.30%. U tabeli 13 su date vrednosti zarada, u Republici Srbiji i susednim zemljama, izražene u domaćoj valuti, evrima i dolarima.

Zemlja	2008	2009	2010	2011	2012
Republika Srbija	32757.17 RSD 402.65 EUR 593.53 USD	31757.83 RSD 338.225 EUR 473.48 USD	34159.25 RSD 332.04 EUR 441.54 USD	38001 RSD 372.94 EUR 520.77 USD	41386 RSD 366.25 EUR 470.30 USD
Republika Hrvatska	5176.83 HRK 717 EUR 1050.07 USD	5311 HRK 723.57 EUR 1005.87 USD	5341.83 HRK 732.76 EUR 971.24 USD	5440.58 HRK 732.24 EUR 1018.84 USD	5477.91 HRK 728.45 EUR 936.39 USD
Mađarska	119297.08 HUF 474.04 EUR 692.22 USD	124385.08 HUF 443.71 EUR 615.31 USD	133754.58 HUF 485.87 EUR 643.11 USD	144504.75 HUF 517.29 EUR 718.64 USD	150741 HUF 516.18 EUR 669.39 USD
Republika Makedonija	16094.75 MKD 262.69 EUR 384.49 USD	19958.33 MKD 325.74 EUR 452.76 USD	20553.67 MKD 334.15 EUR 442.39 USD	20847.83 MKD 338.82 EUR 471.35 USD	20902 MKD 339.48 EUR 436.19 USD
Rumunija	1282.25 RON	1381.08 RON	1406.58 RON	1475.08 RON	1535 RON

	348.43 EUR	326.50 EUR	334.10 EUR	347.90 EUR	344.94 EUR
	508.83 USD	452.81 USD	442.32 USD	483.63 USD	442.36 USD
Republika Bugarska	516.5 BGN 263.52 EUR 385.45 USD	585.08 BGN 298.51 EUR 414.95 USD	641.67 BGN 327.38 EUR 433.56 USD	697.67 BGN 355.95 EUR 494.80 USD	757.83 BGN 386.65 EUR 498.57 USD
Republika Slovenija	902.84 EUR	930.16 EUR	966.63 EUR	987.34 EUR	991.47 EUR

Tabela 13. Zarade u Republici Srbiji i susedstvu, iskazane u evrima i dolarima

Na slici 34 prikazan je trend promene zarada u Srbiji, izražene u evrima.

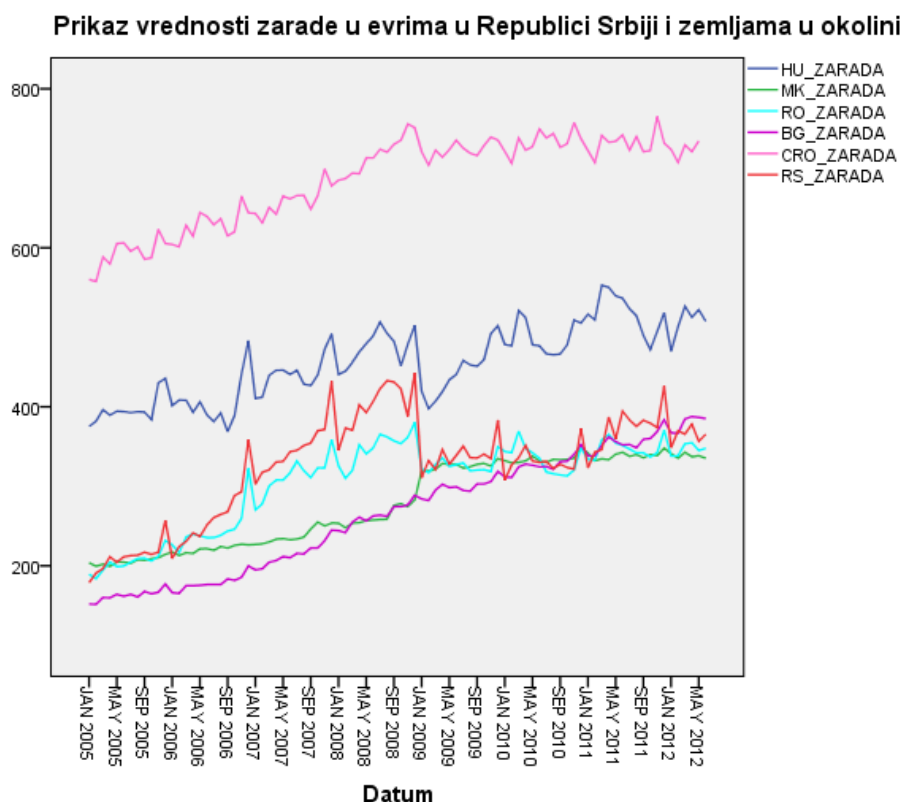


Slika 34. Trend zarade u Republici Srbiji izražene u evrima

Sa slike 34 se primećuje da su zaposleni od 2003. do 2012. najveću platu imali 2008. kada je prosečna zarada bila oko 442 evra (već je rečeno da je najveća zarada iskazana u dinarima bila u decembru 2011., i ona je u evrima imala vrednost 426 EUR). Nakon tog perioda prosečna zarada kreće naglo da pada sve do danas, do 2013. godine, što je u suprotnosti sa trendom zarade izražene u dinarima, jer kada se pogleda slika 33 vidi se da zarada izražena u dinarima ima neprekidno rastući trend. Na grafiku se

primećuje drastičan pad vrednosti plata od oktobra 2008. do februara 2011. godine, prouzrokovan zamrzavanjem plata u 2009. i 2010. godini.

Na slici 35 je prikazana zarada u Srbiji i susednim zemljama, izražena u evrima. Primećuje se da je Srbija po vrednosti zarade u rang sa Bugarskom, Rumunijom i Makedonijom. Ali, kada je u pitanju promena vrednosti zarada primećuje se isti trend u Srbiji, Rumuniji, Mađarskoj i Hrvatskoj.



Slika 35. Prikaz zarada u evrima u Republici Srbiji i zemljama u okolini

Svako slabljenje dinara dovodi do pada standarda, što se može videti i predstavljanjem plate kroz cene proizvoda koju je zaposleni mogao da kupi sa prosečnom platom u periodu od 2006. do 2008. godine. Zaposleni u Republici Srbiji je u **2008.** godini za platu od 32757.17 dinara mogao da kupi 30 vekni hleba¹⁶ za 1050 dinara, 30l mleka¹⁷ za 1737 dinara, 60l evro dizela za 6642 dinara, 1500 kWh¹⁸ struje za 6483 dinara, ako je uzeo kredit čija je mesečna rata 100 evra plaćao bi 8142 dinara. Kad se ovo sve sabere od plate bi mu ostalo 8703 dinara za koje bi trebalo da kupi hranu, piće, proizvode za ličnu higijenu, da plati komunalne usluge, odlazak na odmor, članarinu za biblioteku, itd. U **2010.** godini za platu od 34159.25 dinara zaposleni je mogao da kupi 30 vekni hleba za 1320 dinara, 30l mleka za 1890 dinara, 60l evro

¹⁶ Cena belog hleba „Sava“

¹⁷ Cena litre mleka mlekare Subotica

¹⁸ Prosečna cena električne energije preuzeta sa sajta Elektroprivrede Srbije (<http://www.eps.rs>)

dizela za 7165 dinara, 1500 kWh za 7215 dinara, ako je uzeo kredit čija je mesečna rata 100 evra plaćao bi 10285 dinara. Od plate bi mu ostalo 6284.25 dinara.

U 2012. godini za platu od 41386 dinara zaposleni je mogao da kupi 30 vekni hleba za 1455 dinara, 30l mleka za 2250 dinara, 60l evro dizela za 9210 dinara, 1500 kWh za 9580 dinara, ako je uzeo kredit čija je mesečna rata 100 evra plaćao bi 11300 dinara. Od plate bi mu ostalo 7591 dinara.

Prema ovim informacijama zaposleni u Srbiji su imali najbolje uslove za život u 2008. godini, iako tako ne deluje kada se pogleda tabela sa platama, jer su veće plate (za 6000 – 9000 dinara veće plate) imali 2011. i 2012. godine.

8.3.4. Analiza sezonske komponente

Nad atributom `neto_zarade` primenjena je sezonska dekompozicija kako bi se utvrdilo da li postoji određeni mesec u godini kada je povećanje ili smanjenje zarade bilo izrazito drugačije u odnosu na druge mesece u godini (Tabela 16).

Seasonal Factors

Series Name: `neto_zarade`

Period	Seasonal Factor (%)
1	92.3
2	97.7
3	98.4
4	103.0
5	98.4
6	100.7
7	101.1
8	99.0
9	98.9
10	98.9
11	98.1
12	113.6

Tabela 14. Sezonski faktor za atribut `neto_zarade`

Iz tabele 14, se primećuje da postoji izražen sezonski faktor u decembru – 113.6%. U ovom periodu zaposleni najčešće dobijaju trinaestu platu za predstojeće praznike, te je zbog toga prosečna mesečna plata za decembar svake godine veće. A najmanja vrednost plate (tj. najmanji sezonski faktor) je u januaru svake godine. Pošto su u 2009. i 2010. godini bile zamrznute plate, da bi analiza sezonske komponente bila ispravna mora se ukloniti taj deo. Na osnovu tabele 15 se donosi sličan zaključak kao i na osnovu tabele 14 da postoji veoma izražen sezonski indeks u januaru.

Seasonal Factors

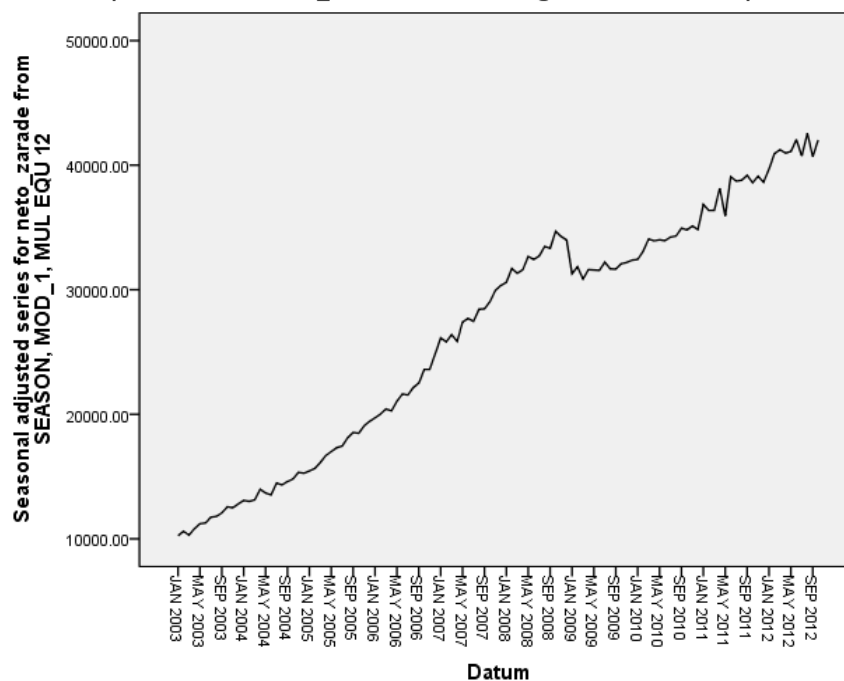
Series Name: neto_zarade

Period	Seasonal Factor (%)
1	92.9
2	98.1
3	98.3
4	103.2
5	98.5
6	101.2
7	100.5
8	99.0
9	98.6
10	98.8
11	97.8
12	113.1

Tabela 15. Sezonski faktor za atribut neto_zarade nakon isključivanja perioda zamrzavanja plate

Zbog izrazitog sezonskog faktora radi se desezoniranje atributa neto_zarade. Na slici 37 je prikazan trend atributa neto_zarade, koji je nazvan SAS_zarade, iz kojeg je uklonjena sezonska komponenta.

Grafički prikaz atributa *neto_zarade* oslobođenog od sezonske komponente



Slika 36. Grafički prikaz atributa *SAS_zarade*

U daljem radu sa podacima o neto zaradi koristi se atribut *SAS_zarade*, jer da bi kasnija istraživanja bila ispravna podaci moraju biti očišćeni od sezonskih faktora.

8.4. Karakteristike atributa *iznos_penzije*

Za analizu promene vrednosti penzija u Republici Srbiji koristi se atribut *iznos_penzije* iz radne sveske PENZIJA.

8.4.1. Deskriptivna statistika

U tabeli 16 je predstavljena deskriptivna statistika atributa *iznos_penzije*.

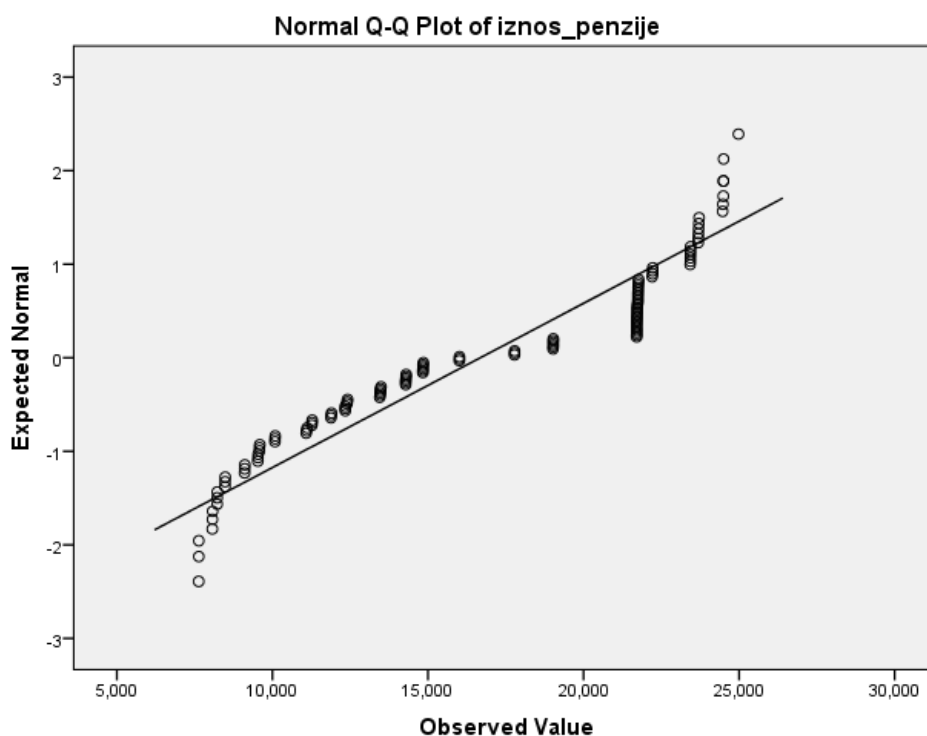
	N	Minimum	Maximum	Mean	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
<i>iznos_penzije</i>	118	7629.15	24979.00	16685.2356	-.125	.223	-1.528	.442
Valid N	118							

Tabela 16. Deskriptivna statistika atributa *iznos_penzije*

Iz tabele 16 se vidi da je u periodu od 2003. do 2012. godine najmanji iznos prosečne penzije bio 7629.15 RSD (u aprilu 2003.), a najveći 24979 RSD (u oktobru 2012.). Prosečna vrednost prosečne penzije u Srbiji u periodu od 2003. do 2012. je 16685.24.

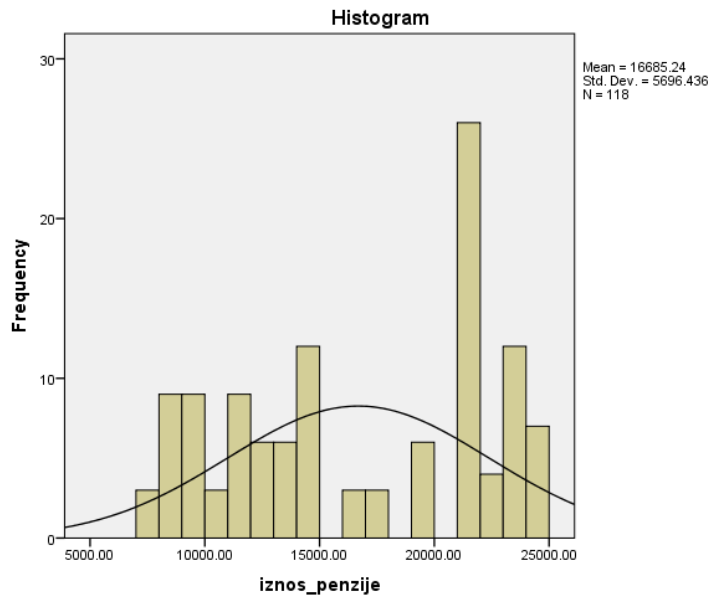
8.4.2. Test normalnosti za atribut iznos_penzije

Uzimaju se informacije o koeficijentima asimetrije i spljoštenosti iz tabele 16. Primenom statističkog testa na koeficijent asimetrije dobija se vrednost: $0.125/0.223=1.112$, što znači da se dobijena vrednost koeficijenta asimetrije ne može uzeti u obzir pri testiranju normalnosti atributa. Zato se radi druga provera preko grafika Normal Q-Q plot. Pošto se vrednosti atributa iznos_penzije ne poklapaju sa regresionom linijom zaključuje se da vrednosti atributa iznos_penzije nemaju normalnu raspodelu (Slika 37).

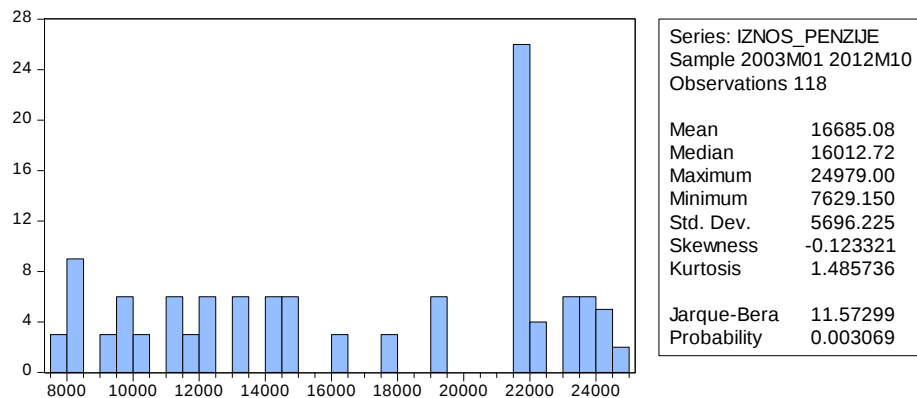


Slika 37. Normal Q-Q Plot za atribut iznos_penzije

Krivulja normalnosti na histogramu govori isto (Slika 38) a i Harke-Berin test (Slika 39).



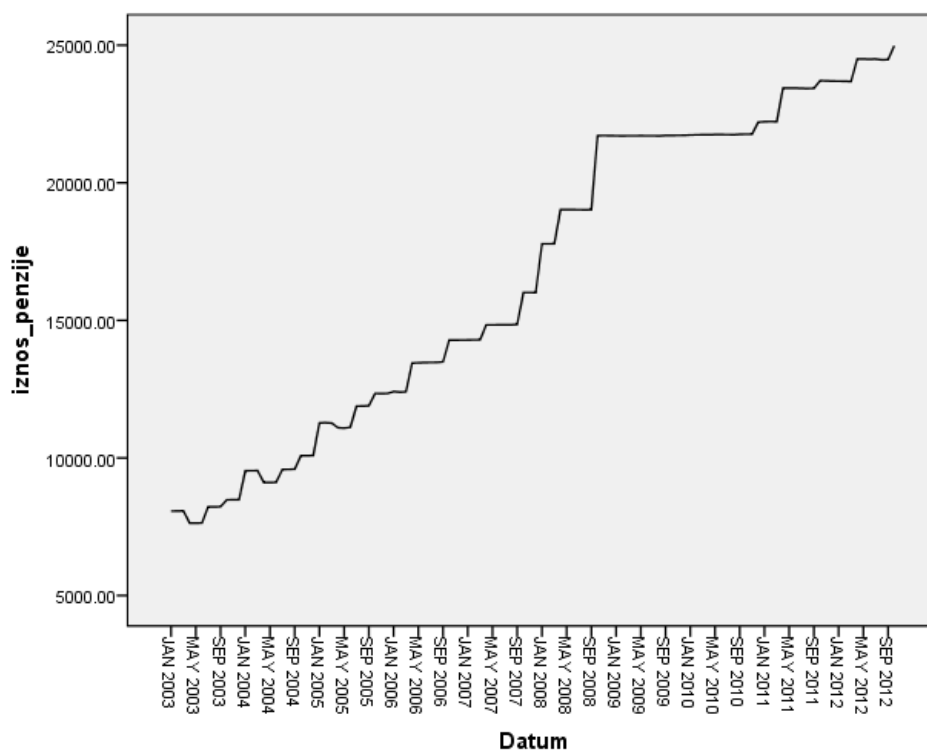
Slika 38. Histogram za atribut iznos_penzije



Slika 39. Rezultat testa Jarque-Bera za atribut iznos_penzije

8.4.3. Analiza trenda

Sa slike 40 se vidi da je vrednost penzije u Republici Srbiji imala rastući trend u periodu od 2003. do 2012. godine. Na grafiku se može uočiti i to da od novembra 2008. do januara 2011. godine iznos penzije prate veoma male promene (promene u vrednosti od 10 dinara), što je rezultat odluke Vlade o zamrzavanju penzija u 2009. i 2010. godini.



Slika 40. Trend penzija u RS

Da bi se izveo zaključak o tome koliko se brzo menjala vrednost penzije u Srbiji, izvršeno je poređenje sa vrednostima penzija koju imaju penzioneri u susednim državama. U tabeli17 su date vrednosti prosečnih godišnjih penzija u Srbiji i zemljama u susedstvu u periodu od 2008. do 2012. godine.

Zemlja	2008	2009	2010	2011	2012
Srbija	19531.66 RSD 239.89 EUR 350.66 USD	21714.35 RSD 231.32 EUR 321.84 USD	21789.93 RSD 211.86 EUR 280.40 USD	23200.17 RSD 227.59 EUR 316.42 USD	24408.42 RSD 216.00 EUR 277.37 USD
Hrvatska	2059.52 HRK 285.25 EUR 417.75 USD	2156.83 HRK 293.85 EUR 408.49 USD	2165.3 HRK 297.02 EUR 393.69 USD	2156.83 HRK 290.13 EUR 403.9 USD	2165.64 HRK 287.99 EUR 370.19 USD
Mađarska	84306 HUF 335.00 EUR 489.18 USD	83393 HUF 297.49 EUR 412.53 USD	86361 HUF 313.70 EUR 415.42 USD	91292 HUF 326.80 EUR 454 USD	93000 HUF 318.46 EUR 413.3 USD
Makedonija	9541 MKD 155.72 EUR	10057 MKD 164.14 EUR	10273 MKD 167.01 EUR	10682 MKD 173.63 EUR	10870 MKD 176.55 EUR

	227.93 USD	233.05 USD	229.92 USD	241.51 USD	226.84 USD
Rumunija	573 RON 155.71 EUR 227.38 USD	686 RON 162.17 EUR 224.92 USD	716 RON 170.07 EUR 225.16 USD	753 RON 178.01 EUR 246.89 USD	768.6 RON 172.72 EUR 246.89 USD
Bugarska	208.97 BGN 56.79 EUR 106.62 USD	244.46 BGN 57.79 EUR 124.72 USD	263.44 BGN 62.72 EUR 134.41 USD	266.63 BGN 63.00 EUR 136.04 USD	Podaci nisu dostupni
Slovenija	554.16 EUR	570.33 EUR	576.88 EUR	578.44 EUR	565.15 EUR

Tabela 17. Prosečne penzije iskazane u domaćoj valuti, evrima i dolarima u Republici Srbiji i zemljama u okolini

Iako naizgled, kada se gledaju penzije izražene u dinarima, deluje kao da se vrednost penzije iz godine u godinu povećava, to nije sasvim tačno. Ako se pogledaju vrednosti penzije iskazane u evrima i dolarima primećuje se da se vrednost penzije ne povećava, već smanjuje. To se najbrže uočava ako se pogledaju penzije iz 2008., 2010. i 2012. godine. 2008. godine je penzija bila 19531 dinara, što je za 2258.27 dinara manje nego 2010. godine i za 4877.42 dinara manje nego 2012. godine. Međutim, ako se pogledaju te iste penzije iskazane u evrima penzija iz 2008. je za 28 evra veća u odnosu na penziju iz 2010. godine i 23 evra veća u odnosu na penziju iz 2012. godine. Slična, pa čak i lošija, situacija je sa penzijom izraženom u dolarima - penzija se od 2008. do 2010. smanjila za 73 dolara, a do 2012. za 97.37 dolara. Takođe, ako se pogleda tabela 9 uočava se da je dinar od 2008. – 2010. u odnosu na evro (dolar) depresirao za 29.92% (39.87%), dok se iz tabele 17 vidi da je penzija u tom periodu porasla za samo 11.56%. Ista je situacija i za period od 2008. – 2012. godine, kada je dinar depresirao za 42.54% (63.41%), a penzija porasla za samo 24.97%. To znači da rast penzije ne prati rast deviznog kursa.

Iz tabele 17 se vidi i da je Srbija 2008. godine imala najveću penziju izraženu u evrima i dolarima. Od te godine vrednost penzije se nije povećavala. I u Mađarskoj je ista situacija. Prema tabeli 17, prosečna penzija u Srbiji je u rangi sa prosečnim penzijama u Hrvatskoj, Makedoniji i Rumuniji, dok Slovenija i Mađarska imaju znatno veće vrednosti prosečne penzije.

Slične informacije o poziciji penzije u Srbiji u odnosu na penzije drugih zemalja se dobija i iz tabele 18. U tabeli 18 su dati procenti promene penzija iskazanih u domaćoj valuti, evrima i dolarima u Srbiji i susednim zemljama.

Zemlja	2008. – 2010.	2010. – 2012.	2008. – 2012.
Srbija	11.56% u RSD		

	-11.68% u EUR -20.04% u USD	12.02% u RSD 1.95% u EUR -1.08% u USD	24.97% u RSD -9.96% u EUR -20% u USD
Hrvatska	5.14% u HRK 4.13% u EUR -5.76% u USD	0.02% u HRK -3.04% u EUR -5.97% u USD	5.15% u HRK 0.96% u EUR -11.38% u USD
Mađarska	2.44% u HUF -6.36% u EUR -15.08% u USD	7.69% u HUF 1.52% u EUR -0.51% u USD	10.31% u HUF -4.94% u EUR -15.51% u USD
Makedonija	7.67% u MKD 7.25% u EUR 0.87% u USD	5.81% u MKD 5.71% u EUR -1.34% u USD	13.93% u MKD 13.38% u EUR -0.48% u USD
Rumunija	24.96% u RON 9.22% u EUR -0.98% u USD	7.35% u RON 1.56% u EUR 9.65% u USD	34.14% u RON 10.92% u EUR 8.58% u USD
Slovenija	4.10% u EUR	-2.03% u EUR	1.98% u EUR

Tabela 18. Procenat promene prosečnih penzija u Republici Srbiji i susednim zemljama

Iz tabele 18 se vidi da su Srbija i Rumunija imale najveći rast penzija izraženih u domaćim valutama. Međutim, ako se pogledaju penzije izražene u evrima i dolarima, vidi se da se penzija u Srbiji iz godine u godinu smanjivala.

8.4.4. Analiza sezone komponente

Nad atributom iznos_penzije primenjena je sezonska dekompozicija kako bi se utvrdilo da li postoji izražen sezonski faktor koji na njega utiče (Tabela 19).

Seasonal Factors

Series Name:

iznos_penzije

Period	Seasonal Factor (%)
1	101.8
2	100.8
3	99.8
4	101.2

5	100.4
6	99.5
7	99.9
8	98.9
9	98.1
10	100.9
11	99.7
12	99.1

Tabela 19. Sezonski faktor za atribut iznos_penzije

Iz tabele 19 se vidi da su najveći iznosi penzije od 2003. do 2012. bili tokom januara i aprila, a najmanji tokom avgusta i septembra. Pošto su u 2009. i 2010. godini bile zamrznute penzije, da bi analiza sezonske komponente bila ispravna mora se ukloniti taj deo. Na osnovu tabele 20 se donosi sličan zaključak kao i na osnovu tabele 19 da je najmanja penzija u septembru, a najveća u januaru.

Seasonal Factors

Series Name:

iznos_penzije

Period	Seasonal Factor (%)
1	102.4
2	101.1
3	99.7
4	101.6
5	100.6
6	99.4
7	99.8
8	98.5
9	97.3
10	101.1
11	99.8
12	98.6

Tabela 20. Sezonski faktor za atribut iznos_penzije bez 2009. i 2010. godine

Pošto nijedan od ovih sezonskih faktora nije izrazito visok/nizak, nema potrebe raditi desezoniranje.

9. Rezultati ispitivanja korelacije atributa

U ovom odeljku su predstavljeni rezultati analize povezanosti atributa.

9.1. Analiza korelacije deviznog kursa EUR_RSD i prosečne neto zarade u periodu od 2003. do 2012. godine

H0: Ne postoji korelacija između promene vrednosti deviznog kursa EUR_RSD i prosečne neto zarade¹⁹.

H1: Postoji korelacija između promene vrednosti deviznog kursa EUR_RSD i prosečne neto zarade.

Pošto nisu ispunjeni uslovi za primenu Pirsonove korelacije, jer samo vrednosti atributa EUR_RSD imaju normalnu raspodelu, za ispitivanje korelacije će se primeniti Spirmanova korelacija ranga. U tabeli 21 prikazani su rezultati računanja Spirmanove korelacije ranga za attribute EUR_RSD i SAS_zarade.

Correlations			Seasonal adjusted series for neto zarade	EUR_RSD
Spearman's rho	Seasonal adjusted series for neto_zarade	Correlation Coefficient	1.000	.865**
		Sig. (2-tailed)	.	.000
		N	118	118
	EUR_RSD	Correlation Coefficient	.865**	1.000
		Sig. (2-tailed)	.000	.
		N	118	118

** . Correlation is significant at the 0.01 level (2-tailed).

Tabela 21. Rezultat primene Spirmanove korelacije ranga na attribute EUR_RSD i SAS_zarade

Spirmanova korelacija ranga ukazuje na to da između atributa EUR_RSD i SAS_zarade postoji veoma jaka pozitivna korelacija koja je statistički značajna ($r_s(116) = .865, p = .000$). Prema tome, donosi se zaključak da između promene vrednosti koju dinar ima za jedan evro i prosečne zarade postoji statistički značajna pozitivna korelacija. Ovo znači da kada dinar depresira (slabi) u odnosu na evro, prosečna zarada raste. I nakon isključivanja perioda kada su plate bile zamrznute iz analize, zaključak ostaje isti. '

¹⁹ Prosečne neto zarade oslobođene od sezonske komponente

9.2. Analiza korelacije deviznog kursa EUR_RSD i prosečne penzije u periodu od 2003. do 2012. godine

H0: Ne postoji korelacija između promene vrednosti deviznog kursa EUR_RSD i prosečne penzije.

H1: Postoji korelacija između promene vrednosti deviznog kursa EUR_RSD i prosečne penzije.

Za ispitivanje korelacije će se primeniti Spirmanova korelacija ranga, jer vrednosti atributa iznos_penzije nemaju normalnu raspodelu. U tabeli 22 su prikazani rezultati računanja Spirmanove korelacije ranga za attribute EUR_RSD i iznos_penzije.

Correlations			EUR_RSD	iznos_penzije
Spearman's rho	EUR_RSD	Correlation Coefficient	1.000	.896**
		Sig. (2-tailed)	.	.000
		N	118	118
	iznos_penzije	Correlation Coefficient	.896**	1.000
		Sig. (2-tailed)	.000	.
		N	118	118

** . Correlation is significant at the 0.01 level (2-tailed).

Tabela 22. Rezultat primene Spirmanove korelacije ranga na attribute EUR_RSD i iznos_penzije

Spirmanova korelacija ranga ukazuje na to da između atributa EUR_RSD i iznos_penzije postoji veoma jaka pozitivna korelacija koja je statistički značajna ($r_s(116) = .896, p = .000$). Prema tome, donosi se zaključak da između promene vrednosti koju dinar ima za jedan evro i prosečne penzije postoji statistički značajna korelacija. Ovo znači da kada dinar depresira u odnosu na evro, prosečna penzija raste.

I nakon isključivanja perioda kada su penzije bile zamrznute iz analize, zaključak ostaje isti.

9.3. Analiza korelacije deviznog kursa USD_RSD i prosečne neto zarade u periodu od 2003. do 2012. godine

H0: Ne postoji korelacija između promene vrednosti deviznog kursa USD_RSD i prosečne neto zarade.

H1: Postoji korelacija između promene vrednosti deviznog kursa USD_RSD i prosečne neto zarade.

I u ovom slučaju se primenjuje Spirmanova korelacija ranga, jer ni vrednosti atributa SAS_zarade, a ni vrednosti atributa USD_RSD nemaju normalnu raspodelu. U tabeli 23 prikazani su rezultati računanja Spirmanove korelacije ranga nad atributima USD_RSD i SAS_zarade.

Correlations

			Seasonal adjusted series for neto_zarade	USD_RSD
Spearman's rho	Seasonal adjusted series for neto_zarade	Correlation Coefficient	1.000	.633**
		Sig. (2-tailed)	.	.000
		N	118	118
	USD_RSD	Correlation Coefficient	.633**	1.000
		Sig. (2-tailed)	.000	.
		N	118	118

** . Correlation is significant at the 0.01 level (2-tailed).

Tabela 23. Rezultat primene Spirmanove korelacije ranga na atribut USD_RSD i SAS_zarade

Spirmanova korelacija ranga ukazuje na to da između atributa USD_RSD i SAS_zarade postoji jaka pozitivna korelacija koja je statistički značajna ($r_s(116) = .633, p = .000$). Prema tome, nulta hipoteza se odbacuje. Na osnovu ovoga se zaključuje da postoji pozitivna korelacija između promene vrednosti koju dinar ima za 1USD i prosečne neto zarade (oslobođene od sezonske komponente). Ovo znači da kada dinar depresira u odnosu na dolar, prosečna zarada raste.

9.4. Analiza korelacije deviznog kursa USD_RSD i prosečne neto zarade u periodu od 2003. – 2009. i u periodu od 2011. – 2012. godine

H0: Ne postoji korelacija između promene vrednosti deviznog kursa USD_RSD i prosečne neto zarade

H1: Postoji korelacija između promene vrednosti deviznog kursa USD_RSD i prosečne neto zarade.

U tabeli 24 su prikazani rezultati računanja Spirmanove korelacije ranga za atribut USD_RSD i iznos_penzije kada se isključi period zamrzavanja zarada.

Correlations

		neto zarade	USD RSD
Spearman's rho	neto_zarade	Correlation Coefficient	1.000
		Sig. (2-tailed)	.
		N	94
	USD_RSD	Correlation Coefficient	.496**
		Sig. (2-tailed)	.000
		N	94

** . Correlation is significant at the 0.01 level (2-tailed).

Tabela 24. Rezultat primene Spirmanove korelacije ranga na atribute USD_RSD i neto_zarade u slučaju kada je iz analize isključen period zamrzavanja zarada (od 2009. - 2011. godine)

Nakon izbacivanju perioda zamrzavanja zarada došlo je do manje promene. Spirmanova korelacija ranga sada ukazuje na to da između atributa USD_RSD i neto_zarade postoji umerena pozitivna korelacija koja je statistički značajna ($r_s(116) = .496, p = .000$). I dalje važi zaključak da postoji korelacija između promene vrednosti koju RSD ima za jedan USD i promene vrednosti prosečne zarade, ali je slabijeg intenziteta.

Ovo znači da su 2009. i 2010. godina značajne pri oceni korelacije zarade i deviznog kursa dolara u periodu od 2003. do 2012. godine.

9.5. Analiza korelacije deviznog kursa USD_RSD i prosečne penzije u periodu od 2003. do 2012. godine

H0: Ne postoji korelacija između promene vrednosti deviznog kursa USD_RSD i prosečne penzije.

H1: Postoji korelacija između promene vrednosti deviznog kursa USD_RSD i prosečne penzije.

U tabeli 25 su prikazani rezultati računanja Spirmanove korelacije ranga za attribute USD_RSD i iznos_penzije.

Correlations			
		iznos_penzije	USD_RSD
Spearman's rho	Correlation Coefficient	1.000	.673**
	iznos_penzije Sig. (2-tailed)	.	.000
	N	118	118
	Correlation Coefficient	.673**	1.000
	USD_RSD Sig. (2-tailed)	.000	.
	N	118	118

** . Correlation is significant at the 0.01 level (2-tailed).

Tabela 25. Rezultat primene Spirmanove korelacije ranga na atribute USD_RSD i iznos_penzije

Spirmanova korelacija ranga ukazuje na to da između atributa USD_RSD i iznos_penzije postoji jaka pozitivna korelacija koja je statistički značajna ($r_s(116) = .673, p = .000$). Prema tome, nulta hipoteza se odbacuje, što dovodi do zaključka da postoji pozitivna korelacija između promene vrednosti koju RSD ima za jedan USD i promene vrednosti prosečne penzije.

9.6. Analiza korelacije deviznog kursa USD_RSD i prosečne penzije u periodu od 2003. – 2009. i u periodu od 2011. – 2012. godine

H0: Ne postoji korelacija između promene vrednosti deviznog kursa USD_RSD i prosečne penzije.

H1: Postoji korelacija između promene vrednosti deviznog kursa USD_RSD i prosečne penzije.

U tabeli 26 su prikazani rezultati računanja Spirmanove korelacije ranga za attribute USD_RSD i iznos_penzije kada se isključi period zamrzavanja penzija.

Correlations			
		iznos_penzije	USD_RSD
Spearman's rho	Correlation Coefficient	1.000	.495**
	iznos_penzije Sig. (2-tailed)	.	.000
	N	95	93
	Correlation Coefficient	.495**	1.000
	USD_RSD Sig. (2-tailed)	.000	.
	N	93	93

** . Correlation is significant at the 0.01 level (2-tailed).

Tabela 26. Rezultat primene Spirmanove korelacije ranga na attribute USD_RSD i iznos_penzije u slučaju kada je iz analize isključen period zamrzavanja penzija (od 2009. - 2011. godine)

U ovom slučaju je došlo do određene promene pri izbacivanju perioda zamrzavanja penzija. Spirmanova korelacija ranga sada ukazuje na to da između atributa USD_RSD i iznos_penzije postoji umerena pozitivna korelacija koja je statistički značajna ($r_s(116) = .495, p = .000$). I dalje važi zaključak da postoji korelacija između promene vrednosti koju RSD ima za jedan USD i promene vrednosti prosečne penzije, ali je slabijeg intenziteta.

Ovo znači da su 2009. i 2010. godina značajne pri oceni korelacije penzije i deviznog kursa dolara u periodu od 2003. do 2012. godine.

9.7. Analiza korelacije prosečne neto zarade i broja zaposlenih

H0: Ne postoji korelacija između prosečne neto zarade i broja zaposlenih.

H1: Postoji korelacija između prosečne neto zarade i broja zaposlenih.

Za analizu ove korelacije korišćena je Spirmanova korelacija ranga, jer vrednosti atributa nemaju normalnu raspodelu. U tabeli 27 su prikazani rezultati računanja Spirmanove korelacije ranga za attribute neto_zarade i br_zaposlenih.

Correlations				
			Seasonal adjusted series for neto_zarade	br_zaposlenih
Spearman's rho	Seasonal adjusted series for neto_zarade	Correlation Coefficient	1.000	-.943**
		Sig. (2-tailed)	.	.000
		N	96	96
	br_zaposlenih	Correlation Coefficient	-.943**	1.000
		Sig. (2-tailed)	.000	.
		N	96	96

** . Correlation is significant at the 0.01 level (2-tailed).

Tabela 27. Rezultat primene Spirmanove korelacije ranga na attribute SAS_zarade i br_zaposlenih

Spirmanova korelacija ranga ukazuje na to da između atributa SAS_zarade i br_zaposlenih postoji veoma jaka negativna korelacija koja je statistički značajna ($r_s(94) = -.943$, $p = .000$). Prema tome, nulta hipoteza se odbacuje, što dovodi do zaključka da postoji korelacija između promene vrednosti broja zaposlenih i promene vrednosti prosečne zarade. Ovo znači da kada broj zaposlenih raste, prosečna mesečna zarada opada i obrnuto - kada broj zaposlenih opada, prosečna mesečna zarada raste.

10. Zaključak

U ovom radu je analizirano pitanje da li se zarada i penzija povećavaju tj. smanjuju u isto vreme kada dinar depresira tj. apresira. Ovaj odnos zarada, penzija i deviznog kursa u Republici Srbiji u periodu od 2003. do 2012. godine analiziran je Spirmanovom korelacijom ranga.

Nad podacima je izvršen test normalnosti i analiza trenda podataka, kako bi se odabrala odgovarajuća metoda. Korišćena je Spirmanova korelacija ranga, jer atributi nad kojima se ispitivanje vršilo nisu ispunjavali potrebne uslove za primenu Pirsonovog koeficijenta korelacije i metode kointegracije. Pre primene Spirmanove korelacije ranga, kako bi rezultati bili ispravni, izvršeno je čišćenje i sređivanje preuzetih podataka. Očišćeni su podaci od nedostajućih i netipičnih vrednosti, kao i od sezonskih faktora. Pri analizi trenda deviznih kurseva, penzija i zarada zaključeno je da su se vrednosti penzija i zarada povećavale kada je dinar depresirao (što je kasnije zaključeno i primenom Spirmanove korelacije ranga). Međutim, ta uvećanja nisu bila srazmerna depresijaciji domaće valute, pa je kada su se vrednosti zarade i penzije predstavile u evrima, zaključeno da se zarade i penzije nisu povećavale već smanjivale tokom godina. Takođe, izvršeno je poređenje zarada i penzija u Republici Srbiji i zarada i penzija u zemljama u okolini, na osnovu čega je donet zaključak da su prosečne zarade u Republici Srbiji među najmanjim u regionu. Prosečne zarade u Srbiji su na nivou prosečnih zarada u Bugarskoj, Rumuniji i Makedoniji. Situacija je ista i sa prosečnim penzijama u Srbiji. Prosečna penzija u Srbiji je na nivou prosečnih penzija u Hrvatskoj, Makedoniji i Rumuniji, dok Slovenija i Mađarska imaju znatno veće vrednosti prosečne penzije.

Rezultati analize korelacije prosečne zarade u Republici Srbiji i prosečnog mesečnog deviznog kursa su sumirani u tabeli 28.

Devizni kurs	Koeficijent korelacije	Značaj
EUR_RSD	$r_s(116) = .865$	$p = .000$
USD_RSD	$r_s(116) = .633$	$p = .000$

Tabela 28. Rezultat primene Spirmanove korelacije ranga nad deviznim kursnim i neto zaradama

Donet je zaključak da:

- između deviznog kursa evra i prosečne mesečne neto zarade postoji veoma jaka pozitivna korelacija. To znači da kada dinar depresira tj. apresira u odnosu na evro, povećava se tj. smanjuje vrednost zarade.
- između deviznog kursa dolara i prosečne mesečne neto zarade postoji jaka pozitivna korelacija. To znači da kada dinar depresira tj. apresira u odnosu na dolar, povećava se tj. smanjuje

vrednost zarade. Nakon uklanjanja perioda zamrzavanja penzija iz analize korelacije se dobija zaključak da postoji pozitivna korelacija, ali slabijeg intenziteta

Na osnovu analiza rađenih u poglavlju 8.3.3 donet je zaključak da plate nisu pratile rast deviznog kursa, iako na osnovu Spirmanove korelacije deluje drugačije. Naime, korelacijom je samo utvrđeno da kada raste devizni kurs, raste i zarada, ali ne i da li je zarada porasla za isti procenat kao i devizni kurs.

Rezultati analize korelacije prosečne penzije u Republici Srbiji i prosečnog mesečnog deviznog kursa su sumirani u tabeli 29.

Devizni kurs	Koeficijent korelacije	Značaj
EUR_RSD	$r_s(116) = .896$	$p = .000$
USD_RSD	$r_s(116) = .673$	$p = .000$

Tabela 29. Rezultat primene Spirmanove korelacije ranga nad deviznim kursevima i prosečnim penzijama

Donet je zaključak da:

- između deviznog kursa evra i prosečne penzije postoji veoma jaka pozitivna korelacija. To znači da kada dinar depresira tj. apresira u odnosu na evro, povećava se tj. smanjuje vrednost penzije. Pri uklanjanju perioda zamrzavanja penzija iz analize korelacije se dobija isti zaključak.
- između deviznog kursa evra i prosečne penzije postoji jaka pozitivna korelacija. To znači da kada dinar depresira tj. apresira u odnosu na dolar, povećava se tj. smanjuje vrednost penzije. Nakon uklanjanja perioda zamrzavanja penzija iz analize korelacije se dobija zaključak da postoji pozitivna korelacija, ali slabijeg intenziteta
- Iz navedenog se zaključuje da je povezanost između promene evra i penzija veća.

Pri analizi trenda donet je zaključak i da je promena vrednosti deviznog kursa dolara u EU i dolara u Srbiji negativno korelisana, što znači da kada dolar jača u EU jača i u Srbiji.

Na osnovu analiza rađenih u poglavlju 8.3.3 donet je zaključak da ni penzije, isto kao i zarade, nisu pratile rast deviznog kursa, iako na osnovu Spirmanove korelacije deluje drugačije. Naime, korelacijom je samo utvrđeno da kada raste devizni kurs, raste i penzija, ali ne i da li je penzija porasla za isti procenat kao i devizni kurs.

U radu je obrađeno i pitanje postojanja povezanosti između povećanja broja zaposlenih i smanjenja tj. povećanja zarada. Spirmanova korelacija ranga, koja je korišćena i u ovom slučaju, pokazala je da između atributa SAS_zarade i br_zaposlenih postoji veoma jaka negativna korelacija koja je statistički značajna ($r_s(94) = -.943, p = .000$). Ovo znači da kada broj zaposlenih raste, prosečna mesečna zarada opada i obrnuto - kada broj zaposlenih opada, prosečna mesečna zarada raste. Ako se uzme u obzir i činjenica da je od 2003. do 2012. godine posao izgubilo približno 390.000 zaposlenih, uočava se da je povećanje prosečne plate samo privid. Jer kada su ovi zaposleni dobili otkaz masa zarade se delila sa

manjim brojem zaposlenih, pa je i logično da ona bude veća. Da su ovi zaposleni ostali na svom poslu, vrednost prosečne plate izražene u dinarima bi bila daleko manja.

Šta dalje?

Ovako sređeni i analizirani podaci se mogu iskoristiti za detaljniju analizu i predviđanje novih vrednosti deviznih kurseva, zarade i penzija.

Dodaci

10.1. Dodatak A: Primer određivanja pokretnog proseka

Primer određivanja pokretnog preseka:

Izračunati trogodišnji pokretni prosek za broj stranih turista u Srbiji.

Prvi korak u računanju trogodišnjeg pokretnog proseka jeste određivanje trogodišnjih pokretnih zbirova. Ukupan broj turista za prve tri godine iznosi 1.182.210, za naredne tri je 1.308.210, itd. U drugom koraku je potrebno podeliti svaki trogodišnji pokretni zbir sa 3, da bi se dobili trogodišnji pokretni proseci. Poslednji element ovog trogodišnjeg pokretnog proseka bi se računao po formuli:

$$x_{n-1} = \frac{x_{n-2} + x_{n-1} + x_n}{3} \quad (12)$$

U tabeli 30 vrednost pokretnog zbira i pokretnih proseka se pozicioniraju u redu srednje godine za tu grupu trojki. Otuda i naziv metoda pokretnih sredina.

Napomena: *Ne može se naći tročlani prosek prvog i poslednjeg originalnog podatka.*

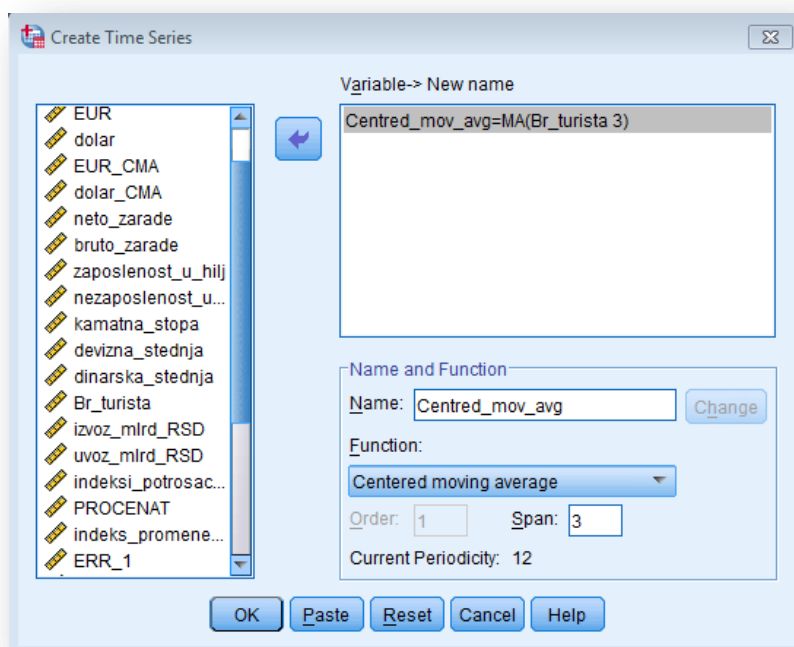
Godina	Broj turista	Tročlani pokretni zbir	Tročlana pokretna sredina	Petočlani pokretni zbir	Petočlana pokretna sredina
2003	340.00				
2004	390.70	1182.21	394.07		
2005	451.51	1308.21	436.07	2344.21	468.84
2006	466.00	1613.51	537.84	2649.21	529.84
2007	696.00	1807.00	602.33	2903.51	580.70
2008	645.00	1986.00	662.00	3135.66	627.13
2009	645.00	1973.67	657.89	3434.23	686.85
2010	683.66	2093.23	697.74	3278.92	707.65
2011	764.56	2248.32	749.41		
2012	800.00				

Tabela 30. Primer računanja pokretnog proseka

Opis primene metode pokretnog proseka u SPSS-u

1. U glavnom meniju odabrati opciju *Transforms* → *Create Time Series*
2. Iz padajućeg menija *Function* odabrati metodu *Centered moving average* i ukucati broj *n* u polje *Span*, čime se definiše da će se računati *n*-člani pokretni prosek.
3. U odsek *Variable* se prevlači numerička kolona koja sadrži vrednosti posmatrane pojave kroz vreme. U ovom primeru to je broj turista. U polju *Name* može se odrediti ime nove promenljive koja će biti kreirana i u koju će biti smešten rezultat tj. pokretni prosek.
4. Klik na dugme *OK*.

Na slici 41 predstavljen je čarobnjak „Create Time Series“.



Slika 41. Create Time Series Wizard

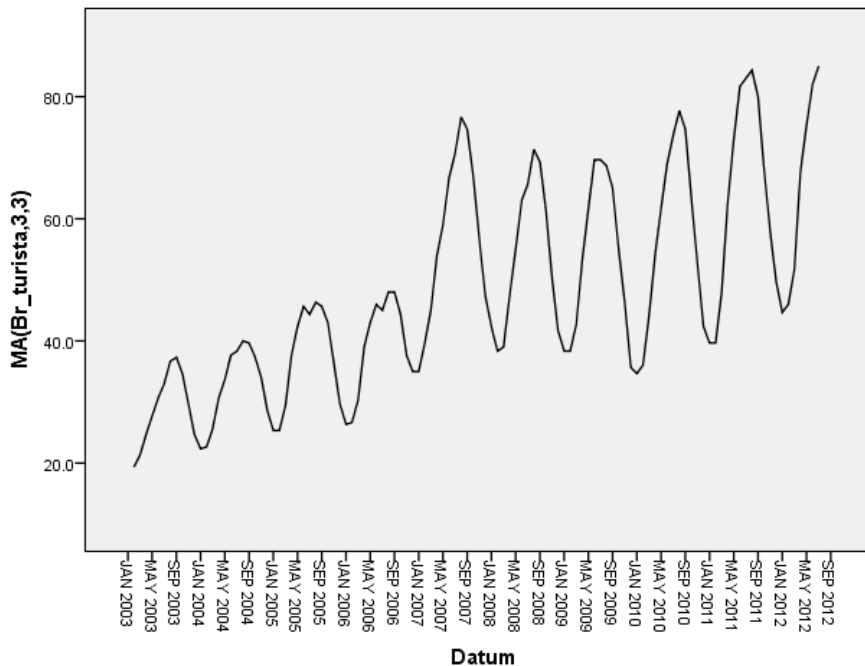
U datoteci sa podacima će se naći nova kolona (u primeru sa slike 41, *Centred_mov_avg*) koja sadrži rezultat. Na slici 42 je prikazana deo nove kolone koja predstavlja tročlanu pokretnu sredinu za atribut broj turista.

Datum	Br_turista	Centred_mov_avg
JAN 2003	18	.
FEB 2003	20	19.3
MAR 2003	20	21.3
APR 2003	24	24.7
MAY 2003	30	27.7
JUN 2003	29	30.7
JUL 2003	33	33.0
AUG 2003	37	36.7
SEP 2003	40	37.3
OCT 2003	35	34.7
NOV 2003	29	29.7
DEC 2003	25	24.7
JAN 2004	20	22.3
FEB 2004	22	22.7
MAR 2004	26	25.7
APR 2004	29	30.7
MAY 2004	37	33.7
JUN 2004	35	37.7
JUL 2004	41	38.3
AUG 2004	39	40.0
SEP 2004	40	39.7
OCT 2004	40	37.3

Slika 42. Nova kolona u datoteci sa podacima

Vrednosti ove nove kolone (Centered_mov_avg) je prikazan na grafiku (Slika 43) kako bi se utvrdio trend.

Prikaz tromesečnog pokretnog proseka za atribut br_turista u RS za period od 2003. do 2012.



Slika 43. Grafički prikaz kolone sa vrednostima tročlanog pokretnog proseka

Trend je rastući, ali to nije tako očigledno. Zbog toga što, kao što se sa grafika vidi, postoji veoma izražena sezonska komponenta. Sezonska komponenta u ovom primeru će kasnije biti uklonjena postupkom sezonske dekompozicije, što se radi da bi se utvrdio pravi trend.

10.2. Dodatak B: Primer određivanja sezonskog indeksa

Primer ručnog određivanja sezonskog indeksa:

Odrediti kvartalni sezonski indeks korišćenjem metode odnosa prema pokretnim prosecima. U tabeli 31 su date vrednosti kvartalne prodaja televizora u milionima od 2000. do 2005. godine.

Godina	Zima	Proleće	Leto	Jesen
2000	6.7	4.6	10.0	12.7
2001	6.5	4.6	9.8	13.6
2002	6.9	5.0	10.4	14.1
2003	7.0	5.5	10.8	15.0
2004	7.1	5.7	11.1	14.5

2005	8.0	.2	11.4	14.9
------	-----	----	------	------

Tabela 31. Kvartalna prodaja televizora od 2000. do 2005.

Računanje kvartalnog sezonskog indeksa se sastoji iz šest koraka [8]:

1. Prvi korak je određivanje četvoročlanog (pošto se radi o kvartalu) pokretnog zbira od 2000. godine. Prva vrednost se računa kao zbir vrednosti kvartala $6.7 + 4.6 + 10.0 + 12.7 = 34.0$. Sledeća vrednost se računa kao zbir $4.6+10.0+12.7+6.5 = 33.8$. Ova procedura se nastavlja za svaki kvartal, sve do 2005. godine. U tabeli 32 je prikazano kako se računa.

Godina	Kvartal	Prodaja	Četvoročlani pokretni zbir	Četvoročlana pokretni proseci	Centralna pokretni proseci	Specifičan sezonski indeks
2000	Zima	6.7				
	Proleće	4.6				
			34.0	8.500		
	Leto	10.0			8.475	1.180
			33.8	8.450		
	Jesen	12.7			8.4	1.503
			33.8	8.450		
2001	Zima	6.5			8.425	.772
			33.6	8.400		
	Proleće	4.6			8.513	0.540
			34.5	8.625		
	Leto	9.8			8.675	1.130
			34.9	8.725		
	Jesen	13.6			8.775	1.50
			35.3	8.825		

Tabela 32. Deo tabele u kojoj su računati četvoročlani pokretni zbir, četvoročlani pokretni prosek, centralni pokretni prosek i specifičan sezonski indeks

2. U drugom koraku se računaju pokretni proseci. U ovom primeru to je četvoročlani pokretni prosek što se dobija tako što se četvoročlani pokretni zbir podeli sa 4. Rezultat se može pogledati u tabeli 32.
3. Formiraju se centrirani pokretni proseci, jer se računa parni pokretni prosek. Prva vrednost se računa kao $(8.500+8.450)/2 = 8.475$, druga vrednost kao $(8.450+8.450)/2 = 8.45$., itd. Rezultat se može pogledati u tabeli 32.
4. Traži se proizvod sezonske i slučajne komponente koji se računa tako što se originalne vrednosti posmatrane pojave (prodaja televizora) podele sa centriranim pokretnim prosecima. Računa se specifični sezonski indeks.
5. U ovom koraku se računa kvartalni prosek da bi se dobio sezonski indeks za određeni kvartal.

Npr., sezonski indeks za leto se dobija tako što se izračuna prosek specifičnih sezonskih indeksa za svako leto od 2000. – 2005. $(1.180+1.130+1.141+1.126+1.143)/5 = 1.144$. Nakon ovog koraka podaci su očišćeni od slučajne komponente.

Radi preglednosti, sve vrednosti specifičnog sezonskog indeksa su prepisane u novu tabelu (Tabela 33).

Godina	Zima	Proleće	Leto	Jesen	
2000			1.180	1.503	
2001	0.772	0.540	1.130	1.550	
2002	0.775	0.553	1.141	1.535	
2003	0.753	0.581	1.126	1.558	
2004	0.733	0.590	1.143	1.466	
2005	0.801	0.615			
Ukupno	3.834	2.879	5.720	7.612	
Prosek	0.767	0.576	1.144	1.522	4.009
Izravdati prosek	0.765	0.575	1.141	1.519	4.000
Indeks	76.5	57.5	114.1	151.9	

Tabela 33. Računanje sezonskog indeksa

6. I u poslednjem koraku računa se korelacioni faktor kako bi se dobio izravdati sezonski indeks. Ovo se radi zbog grešaka u zaokruživanju. Npr. prosek vrednosti (0.767, 0.576, 1.144, i 1.522) jednak je 1.0, te bi njihov zbir trebalo da bude 4.00. Međutim, zbir ovih vrednosti je 4.009, što nije 4.00, već 4.01.

$$\text{Correction factor} = \frac{4.00}{\text{zbir svih tipičnih sezonskih indeksa}} \quad (13)$$

U ovom primeru korelacioni faktor je $\frac{4.00}{4.009} = 0.997755$

Zato će izravdati zimski kvartalni prosek biti $0.767 \cdot 0.997755 = 0.765$, letnja $0.576 \cdot 0.997755 = 0.575$, itd. Svaki od kvartalnih proseka je poravdati nadole kako bi zbir kvartalnih specifičnih sezonskih indeksa bio 4.00. Ako se sada saberu nove izravdate vrednosti, rezultat će biti 4.00.

Pošto se indeksi predstavljaju obično u procentima, svaka dobijena vrednost se množi sa 100. Te je konačna vrednost indeksa za zimski kvartal 76.5, za jesen 151.9,...

Na koji način se dobijene vrednosti interpretiraju?

Recimo, prodaja za jesen je 51.9 procenta iznad prosečnog kvartala, a za zimu 23.5 procenta ispod. Prema tome, u zimskom periodu je poželjno uvesti neke akcije ili povećati broj reklama, kako bi se povećala prodaja u ovom periodu.

Ceo ovaj proces se mnogo brže, jednostavnije i sa manje grešaka može obaviti korišćenjem programa SPSS.

10.3. Dodatak C: Primer desezoniranja

Primer desezoniranja

Na primer, ako pogledamo prethodni primer, desezonirana vrednost za zimu 2000. godine se dobija tako što se sezonski indeks za zimski kvartal (0.765) podeli sa vrednošću prodaje za zimski kvartal 2000. (6.7). Dobijena vrednost je desezonirana vrednost prodaje za zimski kvartal 2000. godine. Ova vrednost sadrži samo trend, ciklus i slučajnu komponentu. U tabeli 34 predstavljene su desezonirane vrednosti pojave iz prethodnog primera.

Godina	Kvartal	Prodaja	Sezonski indeks	Desezonirane vrednosti
2000	Zima	6.0	0.765	8.76
	Proleće	4.6	0.575	8.00
	Leto	10.0	1.141	8.76
	Jesen	12.7	1.519	8.36
2001	Zima	6.5	0.765	8.50
	Proleće	4.6	0.575	8.00
	Leto	9.8	1.141	8.59
	Jesen	13.6	1.519	8.95

Tabela 34. Desezonirana serija

I ovaj proces, isto kao i pronalaženje sezonskog indeksa se može izvršiti preko SPSS-a. Zapravo, određivanje sezonskog indeksa i desezonirana se vrše jednim istim postupkom u SPSS-u – tehnikom *Seasonal Decomposition*.

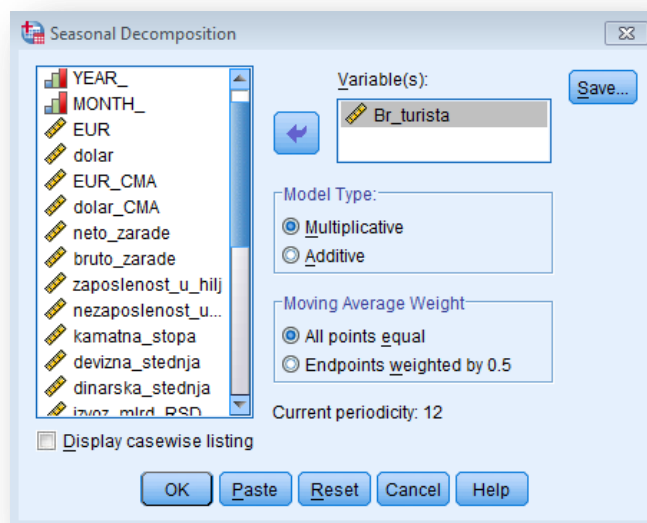
Upotreba SPSS-a za desezoniranje vremenske serije

U SPSS-u postoji procedura koja se zove sezonska dekompozicija i koja omogućava određivanje i uklanjanje sezonske komponente korišćenjem metode odnosa prema pokretnim sredinama. Ova procedura automatski pravi novu seriju u datoteci sa podacima, koja sadrži sezonski poravnate vrednosti (iz 6. koraka), sezonski indeks, trend, cikličnu i slučajnu komponentu.

Koraci za desezoniranje vremenske serije u SPSS-u:

1. U glavnom meniju izabrati opciju *Analyze* → *Forecasting* → *Seasonal Decomposition*.
2. U prozoru Seasonal Decomposition se bira atribut za koji će se raditi desezoniranje. Primera radi izvršiće se desezoniranje atributa broj turista. Opcija *Moving Average Weight* omogućava specifikaciju načina na koji se računaju pokretne sredine. Postoje dva načina:
 - *All points equal* – pokretne sredine se računaju sa korakom jednakim periodičnosti. Ovaj metod se uvek koristi ukoliko je periodičnost neparna.
 - *Endpoint weighted by 5* – pokretna sredina za serije sa parnom periodičnošću se izračunavaju sa korakom jednakim periodičnosti +1 i tada se krajnje tačke računaju sa težinom 0.5.

Na slici 44 predstavljen je prozor Seasonal Decomposition i parametri podešeni za primer desezoniranje atributa broj turista:



Slika 44. Prozor Seasonal Decomposition sa podešenim parametrima

Napomena: Ova metoda neće moći da se primeni, ukoliko u podacima postoje nedostajuće vrednosti.

U tabeli 35 prikazani su sezonski indeksi dobijeni sezonskom dekompozicijom na broj turista u Republici Srbiji:

Seasonal Factors	
Series Name: Br_turista	
Period	Seasonal Factor (%)
1	65.6
2	65.4
3	79.5
4	92.7
5	121.2
6	113.4
7	125.5
8	127.0
9	130.2
10	114.9
11	86.1
12	78.4

Tabela 35. Sezonski indeksi za broj turista

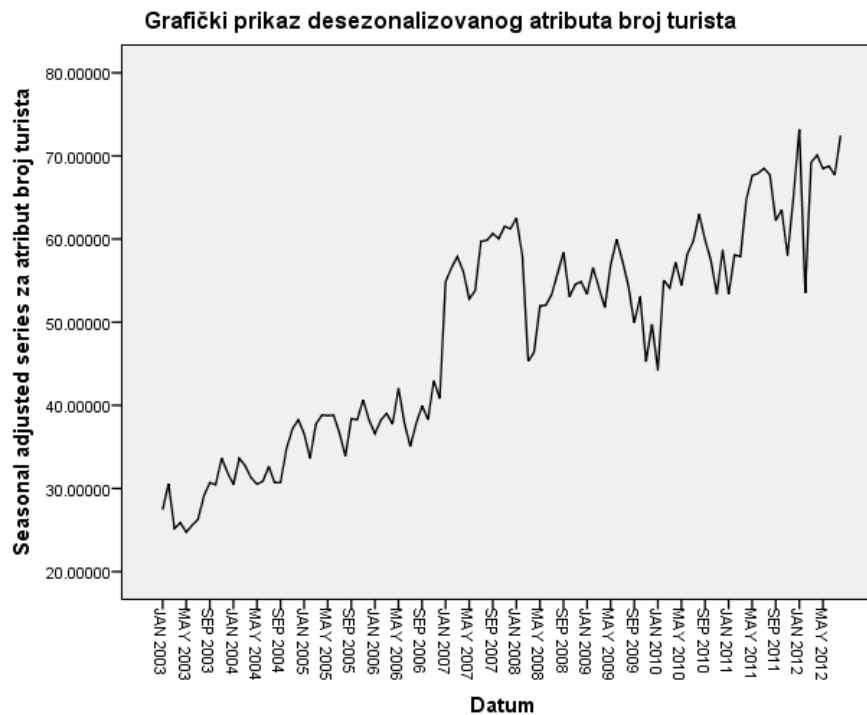
U ovom primeru, indeks 130 za septembar, ukazuje na to da je posećenost turista u septembru bila veća 30% u odnosu na godišnji prosek. A u januaru indeks 65.6, ukazuje na to da je posećenost turista bila manja za 34.4% od godišnjeg proseka. Ova tabela govori da turista ima najviše u julu, avgustu i septembru, a najmanje u januaru i februaru.

U skupu podataka je dodato nekoliko novih kolona koje se kreiraju za svaku originalnu vrednost koja se analizira postupkom sezonske dekompozicije:

- err – predstavlja slučajnu komponenta koja predstavlja vrednost koja ostane kada se uklone sezonska, ciklična i trend komponenta iz serije.
- sas – predstavlja sezonski izravnatu seriju (originalna serija bez sezonske komponente). Ova komponenta omogućava izolaciju sezonske komponente i analizu trenda.
- saf - predstavlja sezonski faktor. Predstavlja vrednost koja ukazuje na efekat svakog perioda na nivo serije. Ako saf ima vrednost 1 (kada se koristi multiplikativni model) to označava odsustvo sezonske komponente, a za aditivni model to je 0.
- stc – predstavlja trend i ciklus, bez sezonske komponente.

Pošto je u Dodatku A donet zaključak (pri analizi trenda atributa Br_turista (Slika 43)) da sezonska komponenta ima veliki uticaj na vrednosti ovog atributa i da zbog toga ne može se sa sigurnošću reći kakav je trend, sada kada je sezonska komponenta otklonjena možemo se vratiti na procenu trenda.

A sa grafika posle uklanjanja sezonske komponente može se zaključiti da je trend rasti (prikazan na slici 45).



Slika 45. Trend desezoniranje vrednosti atributa broj turista

Za X osu je uzet atribut SAS, koji je dodat u skup podataka kao rezultat sezonske dekompozicije, jer on predstavlja originalnu seriju sa uklonjenom sezonskom komponentom. Ovako obrađeni podaci su spremni za predviđanje novih vrednosti ili dalju obradu kako bi se pripremili za analizu međusobnih odnosa atributa što je osnovni zadatak ovog rada.

10.4. Dodatak D: Normalna raspodela

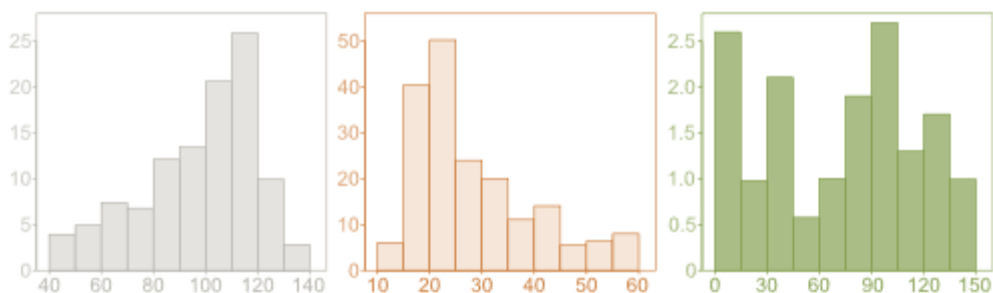
Normalna raspodela

Raspodele podataka se javljaju u različitim oblicima i mogu biti asimetrične (na desnu ili levu stranu) ili simetrične.

Asimetrična raspodela

Asimetrična raspodela je tip raspodele kod koga se aritmetička sredina ne poklapa sa medijanom. Raspodela može biti asimetrična u dva smera: može biti iskrivljena udesno (pozitivno asimetrična) i može biti iskrivljena ulevo (negativno asimetrična). Oblik je pozitivno asimetričan kada je sredina veća od medijane, a negativno asimetričan kada je sredina manja od medijane. Najgori mogući slučaj asimetrične raspodele jeste ukoliko je ona nakrivljena istovremeno i u levo i u desno i ako ima više vrhova (dva vrha – bimodalna, tri vrha tri-modalna skup podataka, ...). Za ovakav tip raspodele kažemo da je izrazito asimetričan.

Na slici 46²⁰, prikazani su primeri histograma asimetrično raspodeljenih podataka.

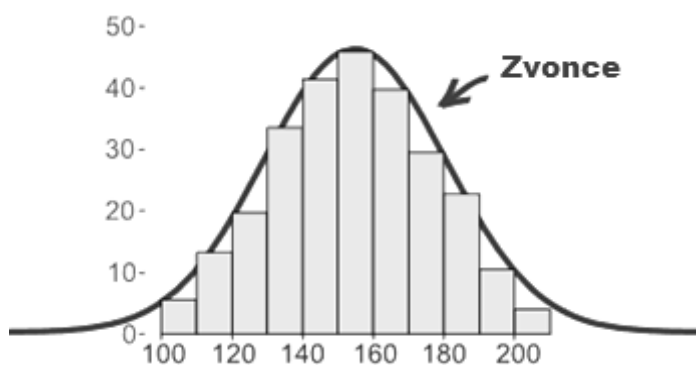


Slika 46. Asimetrično raspodeljeni podaci (pozitivno, negativno i izrazito asimetrični skupovi podataka)

Simetrična raspodela

Raspodela podataka je simetrična ako su joj aritmetička sredina i medijana jednake, i ako joj se krajevi asimptotski približavaju x-osi. Poseban oblik simetrične raspodele je normalna ili Gausova raspodela.

Ako prikupljeni podaci prikazani grafički formiraju simetričnu krivu zvonastog oblika, odnosno ako je većina podataka (vrednosti numeričke varijable x) koncentrisana oko njene sredine dok je mali broj ekstremnih vrednosti smešten na krajevima krive – tada se kaže da vrednosti atributa imaju normalnu raspodelu. Na slici 47²¹ je prikazan primer normalne raspodele.



²⁰ Slika preuzeta sa sajta: <http://www.mathsisfun.com/data/standard-normal-distribution.html>

²¹ Slika preuzeta sa sajta: <http://www.mathsisfun.com/data/standard-normal-distribution.html>

Slika 47. Normalna raspodela

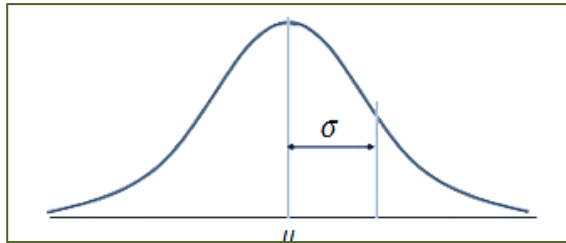
Definicija. Ako je gustina raspodele slučajne veličine X:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in R, \sigma > 0, \mu \in R \quad (14)$$

gde je μ - srednja vrednost, a σ – standardna devijacija

tada se kaže da X ima normalnu raspodelu sa parametrima μ i σ^2 [4].

Prema tome, normalna raspodela zavisi od dva parametra: aritmetičke sredine (μ) i standardne devijacije (σ). Za veće/manje vrednosti standardne devijacije (σ) normalna kriva više je razvučena/sužena, a za manje/veće vrednosti aritmetičke sredine (μ) normalna kriva je spljoštена/zašiljena (Slika 48).

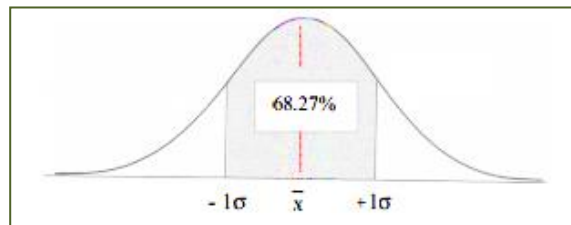


Slika 48. Parametri od kojih normalna distribucija zavisi

68-95-99.7 pravilo (tri sigme ili empirijsko pravilo)

Za normalno distribuirani atribut x važi da se:

- 68.27% njenih vrednosti nalazi unutar intervala $x \pm 1\sigma$ (Slika 49):



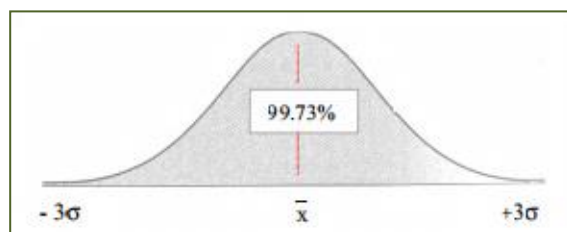
Slika 49. Pravilo 68%

- 95.45% njenih vrednosti nalazi unutar intervala $x \pm 2\sigma$ (Slika 50):



Slika 50. Pravilo 94%

- 99.73% njenih vrednosti nalazi unutar intervala $\bar{x} \pm 3\sigma$ (Slika 51):



Slika 51. Pravilo 99.7%

Test normalnosti

Provera raspodele atributa može se utvrditi na nekoliko načina:

- pomoću koeficijenta asimetrije (eng. *skewness*) i spljoštenosti (eng. *kurtosis*),
- grafika Normal Q-Q plot,
- histograma i krive normalnosti
- Harke-Berinog algoritma
- testovima normalnosti Šapiro-Vilk i Kolmogorov-Smirnov.

Provera raspodele preko koeficijenta asimetrije i spljoštenosti

Koeficijent asimetrije je koeficijent koji ukazuje na to da li je raspodela asimetrična ulevo ili udesno.

Jednak je 0 ako je raspodela simetrična (ako je raspodela savršeno normalna). Može odstupati od 0, sa pozitivnim i negativnim predznakom, ali sve dok je vrednost koeficijenta asimetrije manja od 1, smatra se da raspodela nije asimetrična. Ako je predznak negativan raspodela je nakošena ulevo, tj. u tom slučaju veća je frekvencija natprosečnih vrednosti. Ako je predznak pozitivan, onda je raspodela nakošena

udesno, i to znači da je veća frekvencija ispodprosečnih vrednosti.

Za proveru ispravnost dobijene vrednosti koeficijenta asimetrije radi se statistički test po formuli²²:

$$Z_{g1} = skewness/SES \quad (15)$$

gde je *SES* standardna greška asimetrije koja se računa preko formule:

$$SES = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \quad (16)$$

Da bi se dobijeni koeficijent asimetrije uzeo u obzir kao validan neophodno je da važi:

$$Z_{g1} < -2 \text{ i } Z_{g1} > 2$$

Ovaj test se radi kako bi se utvrdilo da li dobijeni koeficijent asimetrije koji je izračunat na malom uzorku može da se uopšti i da važi za celu populaciju. Što se dobije veći broj, veća je verovatnoća da može.

Koeficijent spljoštenosti je koeficijent koji pruža informaciju o rasprostranjenosti raspodele po y osi, tj. u kojoj su meri vrednosti koncentrisane oko aritmetičke sredine. Koeficijent spljoštenosti ima vrednost 3 za normalnu raspodelu. Tačnije, od vrednosti koja je dobijena za kurtosis treba oduzeti 3 tako da idealna normalna raspodela ima kurtosis = 0. U SPSS-u je ovo urađeno automatski, tako da se odmah prikaže vrednost koja je oduzeta od 3 (što nije slučaj sa EViewsom). Spljoštene raspodele imaju negativnu vrednost, a zašiljene pozitivnu vrednost koeficijenta spljoštenosti. I na koeficijent spljoštenosti se mora primeniti statistički test, kako bi se izmerila validnost dobijenog rezultata:

$$Z_{g2} = kurtosis/SEK \quad (17)$$

gde je *SEK* standardna greška spljoštenosti, koja se računa preko formule:

$$SEK = 2(SES) \sqrt{\frac{n^2 - 1}{(n - 3)(n + 5)}} \quad (18)$$

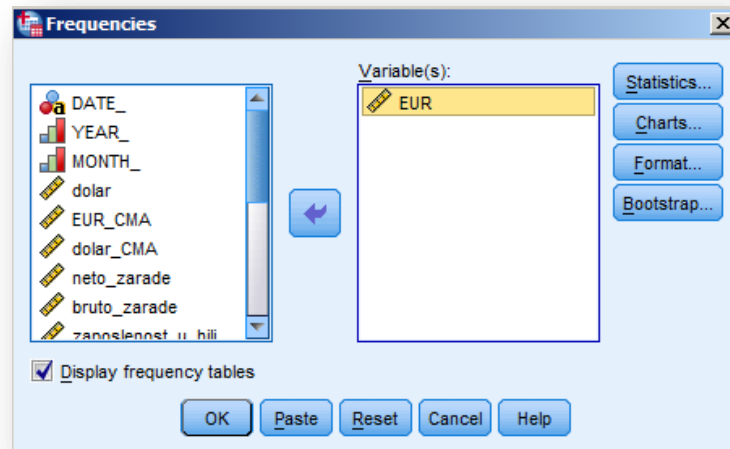
Dobijen koeficijent spljoštenosti se uzima u obzir samo ukoliko je $Z_{g2} < -2$ i $Z_{g2} > 2$

Računanje koeficijenta spljoštenosti i koeficijenta asimetrije preko SPSS-a

Do parametara asimetričnosti i spljoštenost u SPSS-u se dolazi preko opcije deskriptivna statistika (*Analyze* → *Descriptive Statistics* → *Frequencies*). Na slici 52 je prikazan prozor *Frequencies* u kojem je

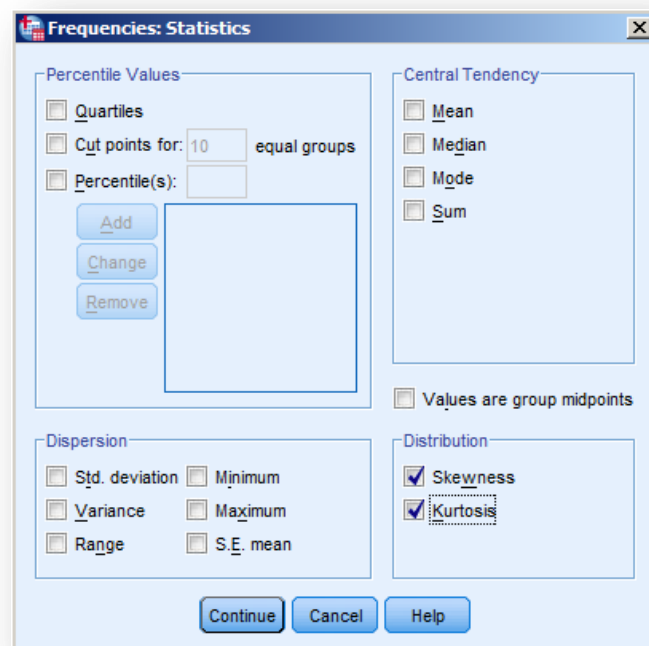
²² U preliminarnim statističkim zaključivanjima se bliskost nuli koeficijenata asimetrije i spljoštenosti dovodi u vezu sa normalnom raspodelom, ali taj zaključak može da bude pogrešan, jer ima i drugih raspodela kod kojih su ovi koeficijenti jednaki 0 [4].

potrebno odabrati atribut za koji se želi napraviti tabela sa deskriptivnom statistikom.



Slika 52. Prozor Frequencies

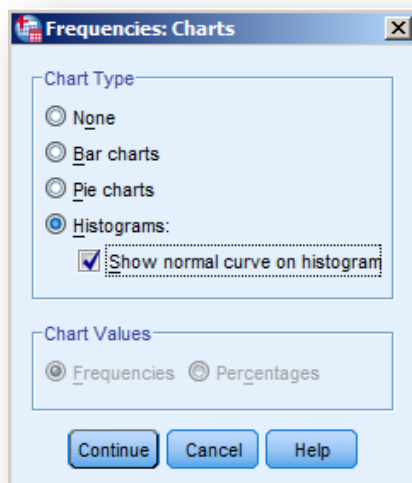
U prozoru koji se dobije kada se klikne na dugme *Statistics* moraju se štiklirati opcije *Skewness* i *Kurtosis* (prikazano na slici 53).



Slika 53. Opcije koje se dobijaju klikom na dugme Statistics

Nakon klika na dugme *Continue* i dugme *OK*, dobija se tabela u kojoj se između ostalog nalaze i vrednosti koeficijenta spljoštenosti i asimetrije.

Za **proveru raspodele preko histograma i krive normalnosti** može se u prozoru koji se dobije kada se klikne na dugme *Charts*, takođe u prozoru *Frequencies* (Slika 53), štiklirati opcija *Histograms*, a onda i opcija „*Show normal curve on histogram*“ (prikazano na slici 54).

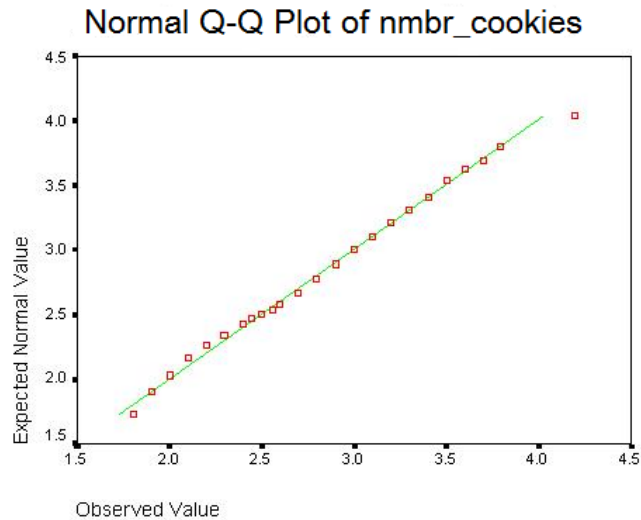


Slika 54. Podešavanje opcije za prikaz histograma

Histogram bi trebalo da ima izgled zvonastog oblika da bi se doneo zaključak da su podaci normalno raspodeljeni.

Provera raspodele preko grafika Normal Q-Q plot

Normal Q-Q plot vrši grafičko poređenje distribucije datog atributa i normalne raspodele koja je na grafiku predstavljena zelenom linijom. Ta linija zove se regresiona linija i prikazuje kako bi grafik izgledao u slučaju da su podaci savršeno normalno raspodeljeni. A crvene tačkice predstavljaju stvarne podatke. Da bi se doneo zaključak da su vrednosti nekog atributa normalno raspodeljene potrebno je da se što više poklapaju sa regresionom linijom. Na slici 55 je prikazana raspodela atributa *nmbr_cookies*, koji ima normalno raspodeljene vrednosti.



Slika 55. Primer provere distribuiranosti preko Normal Q-Q plota

U SPSS-u *Normal Q-Q plot* se dobija preko menija *Graphs* → *Q-Q*.

Provera distribuiranosti preko Harke-Berinog testa

Harke-Berin test je algoritam za testiranje normalnosti koji u testiranje uključuje i parametre asimetrije i spljoštenosti. Računa se preko formule:

$$JB = \frac{n}{6} \left[S^2 + \frac{(k - 3)^2}{4} \right] \quad (19)$$

Gde je:

n broj vrednosti atributa,

S – koeficijenta spljoštenosti,

K – koeficijenta asimetrije.

Ovaj algoritam je podržan u program EViews, preko opcije *Quick* → *Series Statistics* → *Histogram and Stats*. Kada se dobije izlaz najbitniji faktor za proveru distribuiranosti jeste vrednost *Probability*. Ukoliko je *Probability* veća od 5%, onda vrednosti atributa koji se proverava imaju normalnu raspodelu.

10.5. Dodatak E: Primer Spirmanove korelacije ranga

Primer Spirmanove korelacije ranga

1. Rangiranja vrednosti atributa:

Neka su dati atributi težina i godine, prikazani u tabeli 36. Njihovi rangovi su prikazani u trećoj i četvrtoj koloni tabele 36. Najveća vrednost atributa težina (godina) ima rang 1, a najmanja vrednost atributa rang 10 (jer toliko u ovom primeru ima vrednosti).

Težina	Godine	Rang (Težina)	Rang (Godine)
56	66	9	4
75	70	3	2
45	40	10	10
66	60	4	7
61	65	6.5	5
64	56	5	9
58	59	8	8
80	77	1	1
76	67	2	3
61	63	6.5	6

Tabela 36. Rangiranje atributa težina i godine

2. Računa se zbir kvadrata razlike rangova:

$$\sum d_i^2 = 25 + 1 + 9 + 1 + 16 + 1 + 1 = 54$$

Zbrovi kvadrata razlike rangova su predstavljeni u šestoj koloni tabele 37.

Težina	Godine	Rang (Težina)	Rang (Godine)	D	d ²
56	66	9	4	5	25
75	70	3	2	1	1
45	40	10	10	0	0
71	60	4	7	3	9
62	65	6.5	5	1	1
64	56	5	9	4	16
58	59	8	8	0	0
80	77	1	1	0	0
76	67	2	3	1	1
61	63	6.5	6	1	1

Tabela 37. Računanje razlike rangova

I na kraju se računa koeficijent:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} = 1 - \frac{6 * 54}{10 * (10^2 - 1)} = 1 - \frac{324}{990} = 1 - 0.33 = 0.67$$

Pošto je $\rho = 0.67$, zaključak je da je veza pozitivna i jaka.

10.6. Dodatak F: Program XLSManipulation

Program koji popunjava nedostajuće vrednosti u podacima. (Odeljak 7.1)

```
package implementations;

import java.io.FileInputStream;
import java.io.FileNotFoundException;
import java.io.FileOutputStream;
import java.io.IOException;
import java.util.Date;
import org.apache.poi.hssf.usermodel.HSSFCell;
import org.apache.poi.hssf.usermodel.HSSFSheet;
import org.apache.poi.hssf.usermodel.HSSFWorkbook;
import org.apache.poi.ss.usermodel.Row;
/**
 *
 * @author Marijana Lazić
 * 28-10-2012
 * Program koji popunjava nedostajuće datume i vrednosti za te datume. Ima zadatak da prepisuje
 * vrednosti poslednjeg poznatog deviznog kursa na mesta kojima vrednost nedostaje.
 */
public class XLSManipulation {

    public static double findCell(HSSFSheet worksheet, Date cell) {

        /*******Promenljive*****/
        HSSFCell cellA;
        HSSFCell cellB;
        double kurs_stari;
        /*******/

        for (Row row : worksheet) {
            cellA = (HSSFCell) row.getCell((short) 0);
            //Posto je u XLS datoteci kolona C (jer se u njoj nalaze svi datumi, uključujući i vikende) veka od
            //kolone A (u njoj se nalaze datumi bez vikenda), kako ne bismo dobili NullPointerException mora
            //se izvršiti provera da li je vrednost kolone A == null!
```

```

        if (cellA != null) {
            //poredjenje vrednosti kolone A i kolone C koja je prosledjena kao parametar funkcije
            if (cell.equals(cellA.getDateCellValue())) {
                //preuzimanje vrednosti kolone B, koja sadrži podatke o deviznim kursovima, ali bez vikenda
                cellB = (HSSFCell) row.getCell((short) 1);
                kurs_stari = cellB.getNumericCellValue();
                return kurs_stari;
            }
        }
    }
}
return 0;
}

public static void main(String[] args) {

    /*****Promenljive*****/
    FileInputStream fileIN;
    FileOutputStream fileOUT;
    HSSFWorkbook workbook;
    HSSFSheet worksheet;
    double pomocnik1, pomocnik2 = 0, pomocnik3 = 0;
    /*****/

    try {

        fileIN = new FileInputStream("DEVIZNI_KURS.xls");
        fileOUT = new FileOutputStream("DEVIZNI_KURS.xls");
        workbook = new HSSFWorkbook(fileIN);
        worksheet = workbook.getSheet("DEVIZNI_KURS");

        for (Row row : worksheet) {
            HSSFCell cellC = (HSSFCell) row.getCell((short) 3);
            Date datum_novi = cellC.getDateCellValue();
            HSSFCell cellD = (HSSFCell) row.getCell((short) 4);
            if (findCell(worksheet, datum_novi) != 0) {
                //ako findCell vrati vrednost, vrednost nove kolone D se postavlja na vrednost kolone B
                pomocnik1 = findCell(worksheet, datum_novi);
                cellD.setCellValue(pomocnik1);
            } //ukoliko findCell ne vrati vrednost, vrednost kolone D se postavlja na vrednost poslednje
            //pronađene vrednosti
            else {
                if (pomocnik2 != 0) {
                    cellD.setCellValue(pomocnik2);
                    pomocnik3 = pomocnik2;
                } else {
                    cellD.setCellValue(pomocnik3);
                }
            }
        }
    }
}

```

```

    }
    pomocnik2 = findCell(worksheet, datum_novi);
    workbook.write(fileOUT);

    fileIN.close();
    fileOUT.close();
}

} catch (FileNotFoundException ex) {
    System.out.println(ex.getMessage());
}
catch (IOException ex)
{
    System.out.println(ex.getMessage());
}
}
}

```

10.7. Dodatak G: Program HTMLManipulation

Program koji preuzima podatke o izvršenim privatizacijama sa sajta Agencije za privatizaciju i smešta ih u radnu svesku Excel datoteke.

```

package implementations;

import java.io.BufferedReader;
import java.io.FileNotFoundException;
import java.io.FileOutputStream;
import java.io.FileReader;
import java.io.IOException;
import java.util.regex.Matcher;
import java.util.regex.Pattern;
import org.apache.poi.hssf.usermodel.HSSFCell;
import org.apache.poi.hssf.usermodel.HSSFRow;
import org.apache.poi.hssf.usermodel.HSSFSheet;
import org.apache.poi.hssf.usermodel.HSSFWorkbook;

/**
 *
 * @author Marijana Lazić
 * 28-10-2012
 * Program koji preuzima podatke o privatizacijama sa sajta Agencije za privatizaciju.
 * Kada preuzme te podatke, otvara ih i iz njih preuzima datume potpisivanja ugovora
 * o privatizaciji i smešta ih u radnu svesku.
 */

```

```

public class HTMLManipulation {

    public static void wgetcall() throws IOException
    {
        for(int i=1000;i<10120;i++)
        {
            String commandString;
            commandString="wget
http://www.priv.rs/Agencija+za+privatizaciju/90/IRMOVO.shtml/seo=/companyId="+i;
            Process process = Runtime.getRuntime().exec(commandString);
        }
    }
    /**
     * @param args the command line arguments
     */
    public static void main(String[] args) {
        /*****Promenljive*****/
        //String "file" čuva ime datoteke
        String file;
        //radna sveska i list
        HSSFWorkbook myWorkBook;
        HSSFSheet mySheet;
        //čuva izvučeni podatak o datumu
        String rezultat = null;
        //čuva trenutno učitano liniju datoteke
        String line;
        //čuva splitovane podatke linije koja sadrži podatak
        String lines[];
        //čuva šablon koji želimo da se pretraži u datoteci sa pod. o privatizaciji - tražimo datum prodaje
        String pattern_helper;
        //string koji se traži u matcheru
        Pattern regPat;
        //tekst u kojem se proverava da li postoji šablon - regPat2
        Matcher matcher;
        //za čitanje datoteke - privatizacija
        BufferedReader reader;
        //za pisanje rezultata u .xls datoteku
        FileOutputStream fileOUT = null;
        //promenljive za red i kolonu u excel tabeli
        HSSFRow myRow;
        HSSFCell myCell;
        //brojač redova
        int row=0;
        /*****/
        //kreiranje nove radne sveske i lista u Excel dokumentu
        myWorkBook = new HSSFWorkbook();
    }
}

```

```

mySheet = myWorkBook.createSheet();
try {
    //poziv funkciji koja preuzima stranice o privatizovanim preduzećima sa
    //sajta Agencije za privatizaciju. Sve stranice će biti sačuvane u obliku
    //companyid1, companyid2..., companyidn
    wgetcall();
    //otvaramo excel datoteku u koju će se pisati rezultat
    fileOUT = new FileOutputStream("Privatizacija.xls");
    //petlja koja postavlja promenljivu file na ime datoteke koja će se pretraživati.
    //datoteke imaju nazive u obliku companyid1, companyid2..., companyidn
    for (int i = 1000; i < 10120; i++) {

        file = "companyid=" + i;
        //postavljamo vrednost na šablon koji proveravamo
        pattern_helper = "<td class=\"w160\">Datum prodaje</td>";
        regPat = Pattern.compile(pattern_helper);
        //matcher se postavlja na prazan string,
        //kasnije mu se dodeljuje linija po linija datoteke o privat.
        matcher = regPat.matcher("");
        //otvara datoteku za čitanje datuma privatizacije
        reader = new BufferedReader(new FileReader(file));
        //postavljamo liniju na prvu liniju u datoteci
        line = reader.readLine();
        //iščitavanje redova sve dok ne dođemo do kraja datoteke
        while ((line = reader.readLine()) != null) {
            //postavljamo vrednost matchera na prvu liniju datoteke
            matcher.reset(line);
            //ukoliko je pronađen šablon, učitaj novu liniju i
            //izdvoj datum
            if (matcher.find()) {
                //procitaj vrednost
                line = reader.readLine();
                //<td class=\"w160\">15.mart 2004.</td>
                lines = line.split(">");
                line = lines[1];
                lines = line.split("<");
                rezultat = lines[0];
                //ispiši pronađeni datum na standardni izlaz
                System.out.println(rezultat);
                //kada pronađeš taj datum, ne traži dalje.
                //sa tom datotekom je završeno, pređi na sl. datoteku
                break;
            }
        }
        //ukoliko je pronađena vrednost iz prethodne petlje,
        //onda je zapiši u excel radnoj svesci
        if (!"".equals(rezultat)) {

```

```

myRow = mySheet.getRow(row);
//ako red nije kreiran, kreiraj ga
if (myRow == null) {
    myRow = mySheet.createRow(row);
}
myCell = myRow.createCell(0);
myCell.setCellValue(rezultat);
//inicijalizuj na prazan string. Ako sledeća iteracija nema datum
//(tj.ne vraća rezultat), ispisalo bi se prazno polje u excelu.
//upravo zbog toga nakon završenog smeštanja datumu u polje u excelu
//rezultat se postavlja na prazan string.
rezultat = "";
row++;
}
//upiši u radnu svesku
myWorkBook.write(fileOUT);
}
//zatvori radnu svesku
fileOUT.close();
} catch (FileNotFoundException ex){
    System.out.println(ex.getMessage());
} catch (IOException ex) {
    System.out.println(ex.getMessage());
}
}
}
}

```

11. Literatura

- [1] IBM: Dictionary of IBM & Computing Terminology,
<http://www-03.ibm.com/ibm/history/documents/> (posećeno u januaru 2013.)
- [2] Pang-Ning Tan, Michael Steinbach, Vipin Kumar: "Introduction to Data Mining", 2nd Edition, Pearson Education, 2006.
- [3] dr Nenad Mitić, Slajdovi na temu istraživanja podataka, Matematički fakultet, Univerzitet u Beogradu:
<http://poincare.matf.bg.ac.rs/~nenad/ip.html> (posećeno u januaru 2013.)
- [4] Vesna Jevremović: „Verovatnoća i statistika“, Matematički fakultet, Beograd, 2009.
- [5] Julie Pallant: SPSS Survival Manual, 3rd Edition, Allen & Unwin, 2009.
- [6] Eric Zivot: "Introduction to Computational Finance Coursera", University of Washington, 2012/2013.
<https://www.coursera.org> (posećeno u januaru 2013.)
- [7] Kovacic Zlatko: "Time series analysis", Faculty of Economics, University of Belgrade, 1995.
- [8] Prof. dr Mirjana Šekarić: „Statističke metode sa zbirkom zadataka“, Univerzitet Singidunum, Beograd, 2010.
- [9] Jelena Marinković: „Vodiči dobre statističke prakse“, Institut za medicinsku statistiku i informatiku, 2009.
- [10] Stevo Kovačević: Međunarodni ekonomski odnosi, Ekonomski fakultet, Kragujevac, 2000.
- [11] Zakon o radu ("Službeni glasnik RS", br. 24/2005, 61/2005 i 54/2009)
- [12] Grupa autora: „Popularna enciklopedija, BIGZ, Beograd, 1976.
- [13] Zakon o penzijskom i invalidskom osiguranju ("Službeni glasnik Republike Srbije", br. 34/03, 64/04, 84/04, 85/05, 101/05, 63/06, 5/09, 107/09 i 101/10)
- [14] Zakon o budžetu Republike Srbije za 2013. godinu („Službeni glasnik Republike Srbije”, br. 114/12)

[15] Dr Mirko Savić, Slajdovi na temu analize vremenskih serija, Ekonomski fakultet u Subotici, Univerzitet u Novom Sadu: <http://www.ef.uns.ac.rs/Download/statistika> (posećeno u januaru 2013.)