Usporedba grafičkih procedura u prikazu podataka u stočarstvu koristeći SAS program

Papišta, Sven

Master's thesis / Diplomski rad

2019

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Agriculture / Sveučilište u Zagrebu, Agronomski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:204:882502

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Repository / Repozitorij:

Repository Faculty of Agriculture University of Zagreb

zir.nsk.hr

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

Sveučilište u Zagrebu
Agronomski fakultet

University of Zagreb
Faculty of Agriculture

USPOREDBA GRAFIČKIH PROCEDURA U PRIKAZU PODATAKA U STOČARSTVU KORISTEĆI SAS PROGRAM

DIPLOMSKI RAD

Sven Papišta

Zagreb, rujan, 2019.

Sveučilište u Zagrebu
Agronomski fakultet

University of Zagreb
Faculty of Agriculture

Diplomski studij:

Genetika i oplemenjivanje životinja

USPOREDBA GRAFIČKIH PROCEDURA U PRIKAZU PODATAKA U STOČARSTVU KORISTEĆI SAS PROGRAM

DIPLOMSKI RAD

Sven Papišta

Mentor:

Prof. dr. sc. Alen Džidić

Zagreb, rujan, 2019.

Sveučilište u Zagrebu
Agronomski fakultet

University of Zagreb
Faculty of Agriculture

IZJAVA STUDENTA

O AKADEMSKOJ ČESTITOSTI

Ja, **Sven Papišta**, JMBAG 0178100351, rođen 21.09.1995. u Zagrebu, izjavljujem da sam samostalno izradio diplomski rad pod naslovom:

Usporedba grafičkih procedura u prikazu podataka u stočarstvu koristeći SAS program

Svojim potpisom jamčim:

- da sam jedini autor ovoga diplomskog rada;
- da su svi korišteni izvori literature, kako objavljeni tako i neobjavljeni, adekvatno citirani ili parafrazirani, te popisani u literaturi na kraju rada;
- da ovaj diplomski rad ne sadrži dijelove radova predanih na Agronomskom fakultetu ili drugim ustanovama visokog obrazovanja radi završetka sveučilišnog ili stručnog studija;
- da je elektronička verzija ovoga diplomskog rada identična tiskanoj koju je odobrio mentor;
- da sam upoznat s odredbama Etičkog kodeksa Sveučilišta u Zagrebu (Čl. 19).

U Zagrebu, dana _____ _____

Potpis studenta

Sveučilište u Zagrebu
Agronomski fakultet

University of Zagreb
Faculty of Agriculture

IZVJEŠĆE

O OCJENI I OBRANI DIPLOMSKOG RADA

Diplomski rad studenta **Sven Papišta**, JMBAG 0178100351, naslova

Usporedba grafičkih procedura u prikazu podataka u stočarstvu koristeći SAS program

obranjen je i ocijenjen ocjenom _____ , dana _____ .

Povjerenstvo:			potpisi:
1.	Prof. dr. sc. Alen Džidić	mentor	_____
2.	Doc. dr. sc. Dragica Šalamon	član	_____
3.	Doc. dr. sc. Maja Ferenčaković	član	_____

Zahvala

Zahvaljujem se svome mentoru prof. dr. sc. Alenu Džidiću na stručnim savjetima, vodstvu i pomoći koju mi je pružio prilikom izrade ovoga diplomskog rada.

Posebno se želim zahvaliti svojim roditeljima, prijateljima i curi koji su me podupirali tijekom cijelog studija i pomogli da mi vrijeme provedeno na fakultetu bude jedno od najljepših u životu.

Sadržaj

Sažetak *1*

Summary *2*

1. Uvod *3*

2. Opis podataka *4*

3. SGPLOT procedura *5*

3.1. Osnovni SGPLOT grafikoni **6**

3.1.1. Raspršeni grafikon (eng. Scatter plot) 6

3.1.2. Linijski grafikon (eng. Series plot) 8

3.1.3. Mjehuričasti grafikoni (eng. Bubble plot) 11

3.2. **Grafikoni prikaza modela i intervala pouzdanosti** **13**

3.2.1. Regresijski grafikon 13

3.2.2. Grafikon loess funkcije 17

3.3. Distribucijski grafikoni **20**

3.3.1. Histogrami i krivulja funkcija gustoće 20

3.3.2. Krivulja gustoće funkcije (eng. Density function) 22

3.3.3. Pravokutni grafikoni **(eng. Box and Whishers plot)** 24

3.4. Grafikoni distribucija kategoričkih varijabli **28**

3.4.1. Stupčasti grafikon (eng. Bar chart) 28

3.4.2. Točkasti grafikon (eng. Dot plot) 31

4. GPLOT procedura *34*

4.1. Raspršeni grafikon (eng. Scatter plot) **34**

4.2. Mjehuričasti grafikona (eng. Bubble plot) **36**

4.3. Višeslojni grafikon (eng. Overlay plot) **39**

5. Usporedba SGPLOT i GPLOT procedura *42*

6. Zaključak *49*

7. Literatura *50*

Životopis *52*

Sažetak

Diplomskog rada studenta **Svena Papište**, naslova

USPOREDBA GRAFIČKIH PROCEDURA U PRIKAZU PODATAKA U STOČARSTVU KORISTEĆI SAS PROGRAM

U ovom radu uspoređujemo uporabu standardnih GPLOT i naprednih SGPLOT procedura u grafičkom prikazu stočarskih podataka. Grafičke procedure prisutne u SAS programu omogućuju nam efikasan način prikaza podataka. Ovaj rad će prikazati slijedeće tipove grafikona: osnovne grafikone, grafikone koji prikazuju modele s intervalom pouzdanosti, grafikone koji prikazuju distribuciju podataka, te grafikone koji pokazuju kategorizaciju podataka. Dolaskom SAS® 9.2 verzije 2008. godine standardna SAS/GRAPH procedura može se zamijeniti modernijim te za korištenje jednostavnijim SG procedurama.

U ovom radu prikazane su mogućnosti obje procedure, te smo ih grafički prikazali i objasnili. Postoji mnogo prednosti SGPLOT procedure nad GPLOT procedurom. Neke od njih su jednostavnije pisanje koda, bolja kvaliteta grafikona, više ugrađenih opcija i mnoge druge. Navedene prednosti SGPLOT procedure nad GPLOT procedurom u radu smo prikazali i objasnili na primjerima.

Ključne riječi: SAS, SAS/GRAPH, SGPLOT, GPLOT, grafičke procedure

Summary

Of the master's thesis – student **Sven Papišta,** entitled

Comparison of graphical procedures in presenting livestock data using SAS program

In this paper standard GPLOT and advanced SGPLOT graphical procedures for the presentation of livestock data were used. Graphical procedures available in the SAS program are providing us with an effective way of presenting data. This paper will present several graphical procedures: basic graphs, graphs showing models with confidence intervals, graphs showing data distribution, graphs showing data categorization. With arrival of SAS® 9.2 version in 2008 we can replace standard SAS/GRAPH procedures with modern and much simpler to use SG procedures.

In this paper we tested and explained in detail the capabilities of both graphical procedures. SGPLOT procedure has many advantages when compared to the GPLOT procedure. Some of them are easier coding, better graph quality, built in options, and many more. In this paper we tested and explained on examples these specified advantages of SGPLOT procedure over GPLOT.

Keywords: SAS, SAS/GRAPH, SGPLOT, GPLOT, graphical procedures

1. Uvod

Ovaj rad bavi se opisom osnovnih grafikona koji se mogu prikazati koristeći standardnu proceduru GPLOT i naprednu SGPLOT proceduru pomoću SAS programa. Grafikoni koji se mogu prikazati uključuju: osnovne grafikone, grafikone koji prikazuju modele s intervalom pouzdanosti, grafikoni koji prikazuju distribuciju podataka, te grafikoni koji pokazuju kategorizaciju podataka. Rezultate i zaključke statističke analize ili prikaz distribucije stočarskih podataka je važno prikazati na jasan i razumljiv način, te stoga koristimo navedene grafikone. Koristeći podatke iz područja stočarstva (kvantitativne i kvalitativne) usporedit ćemo grafikone napredne grafičke procedure SGPLOT te standardne procedure GPLOT s detaljnim opisom i zaključkom.

Cilj rada je prikazati i usporediti primjenu SGPLOT i GPLOT procedura te opisati njihovu primjenu na podacima iz područja stočarstva. Za obje grafičke procedure objasnit će se značajke sintakse za sve navedene vrste grafikona te na primjeru iz područja stočarstva će se objasniti prednosti i mane navedenih grafičkih procedura.

2. Opis podataka

Za prikazivanje i opisivanje grafičkih procedura u ovome radu korišteno je više setova podataka. Podaci korišteni za izradu ovoga rada su podaci iz godišnjih izvješća Hrvatske poljoprivredne agencije za govedarstvo za razdoblje od 2014. do 2018. godine. Oni su korišteni za prikaz slijedećih grafikona: serijski, mjehuričasti, pravokutni, stupčasti i točkasti.
Za izradu raspršenih, regresijskih linija i Loessove krivulje korišteni su podaci preuzeti s web stranice EUROSTAT-a (https://ec.europa.eu/eurostat/web/products-datasets/product?code=apro_mk_colm).
Podaci iz knjige profesora Kapša „Biostatistics for Animal Science“ su korištenu u izradi histograma i grafikoni gustoće.

Podaci Hrvatske poljoprivredne agencije korišteni u prikazu procedura su godina uzorkovanja (Godina), pasmina (Pasmina), količina mlijeka u kilogramima (MY), broj laktacije u razredima od 1 do 4 (Laktacija), postotak mliječne masti od ukupne količine mlijeka (MM), postotak bjelančevine od ukupne količine mlijeka (Bjelancevine) za Simentalsku, Holstein i Smedju pasminu.

Podaci preuzeti s web stranice Eurostat-a korišteni u prikazu procedura su vrijeme izraženo u mjesecima (TIME), zemlja proizvođač (GEO), vrsta proizvoda (DAIRYPROD), način prikaza podataka (UNIT), vrijednost količine mlijeka u tisućama tonama (Value), postotak mliječne masti od ukupne količine mlijeka (MM (%)), postotak proteina u mlijeku od ukupne količine mlijeka (PROT (%)).

Podaci korišteni iz knjige „Biostatistics for Animal Science“ su težina teladi (Tezina) i razredi (Razred). Podaci o težini teladi u kilogramima pri dobi od 7 mjeseci prikazani su u 15 razreda.

3. SGPLOT procedura

SGPLOT jedna je od tri osnovne statističke grafičke procedure uz SGPANEL i SGSCATTER koje su dostupne od SAS® 9.2 verzije (2008) statističkog programa dolaskom ODS Graphics procedura. Prema Larkhamu (2016) dolaskom SAS® 9.3 verzije (2011) statističkog programa SG procedure su premještene iz paketa SAS/GRAPH u paket Base SAS, tako da su dostupne uz nabavku osnovnog paketa. Nadalje, prema Canchola i sur. (2018) dolaskom SAS® 9.4. verzije (2013) statističkog programa omogućena je lakša automatizacija i prilagodba grafikona.

Prije dolaska posljednjih verzija SAS programa SG se moglo tumačiti kao "Statistical Graphics", ali tijekom godina daljnjim napretkom i unaprjeđenjima SAS SG procedure su postale puno više od "Statistical Graphics". Prema Canchola i sur. (2018) akronim SG može se danas prije protumačiti kao "Splendid Graphics" ili "SAS Graphics". Prema Delwiche i Slaughter (2012) SGPLOT stvara jedan ili više grafikona te ih preklapa na istom koordinatnom sustavu. Prema Matange i Heath (2011) SGPLOT podržava preko dvadeset različitih vrsti naredba za izradu grafikona te naredbi za prilagodbu legendi, osi grafikona i teksta unutar grafikona.

Prema Delwiche i Slaughter (2012) grafikona koji se mogu prikazati SGPLOT procedurom mogu se svrstati u tri osnovne kategorije:

Tablica 1. Kategorizacija grafikona prema Delwiche i Slaughter (2012)

X Y Grafikoni	Trakasti, mjehuričasti, elipse, loess, igličasti, penalizirani B-spline, regresijski, raspršeni, i linijski
Grafikoni distribucija kontinuiranih varijabli	Pravokutni, grafikoni gustoće i histogrami
Grafikoni distribucija kategoričkih varijabli	Točkasti, stupčasti i linijski

Izvor: Delwiche i Slaughter (2012)

Unutar kategorije moguće je preklapati više grafikona ukoliko to ima smisla s obzirom na podatke koje želimo prikazati (Delwiche i Slaughter, 2012).

Sintaksa SAS programa za SGPLOT proceduru:

1. PROC SGPLOT < DATA = set podataka> < opcije >;
2. Plot-naredba(e) potrebni parametri < / opcije>;
3. Podržane naredbe

RUN;

Prema Matange i Heath (2011) za prikaz podataka koristi se jedna ili više plot naredbi, te svaka naredba sadrži vlastite opcije za prikaz podataka. Plot naredbe mogu se spojiti u jednom koraku procedure ukoliko se sa nekoliko grafikona mogu prikazati podaci te ih možemo grupirati kao:

- Osnovni grafikoni: raspršeni, linijski i mjehuričasti grafikoni
- Grafikoni prikaza modela i intervala pouzdanosti: regresijski i linijski grafikoni gdje su točke povezane krivuljom
- Distribucijski grafikoni: histogrami i pravokutni grafikoni
- Kategorizacijski grafikoni: Stupčasti i točkasti grafikoni

Isti autori navode podržane naredbe koje koristimo za prilagodbu izgleda grafikona:

- Referentne linije, tekst unutar grafikona, osi i legende.

3.1. Osnovni SGPLOT grafikoni

3.1.1. Raspršeni grafikon (eng. Scatter plot)

Delwiche i Slaughter (2012) tvrde da su raspršeni grafikoni sjajan način za prikaz veze između dvije kontinuirane varijable, nezavisne varijable koja se nalazi na x osi, i zavisne varijable koja se nalazi na y osi. Prema Matange i Heath (2011) prilikom prikaza podataka pomoću raspršenog grafikona može se promijeniti transparentnost točaka čime se omogućuje bolji pogled na regije veće gustoće podataka. Raspršeni grafikoni jedan su od glavnih načina za vizualizaciju neobrađenih podataka (Delwiche i Slaughter, 2012).

Sintaksa SAS programa prema Matange i Heath (2011):

PROC SGPLOT;

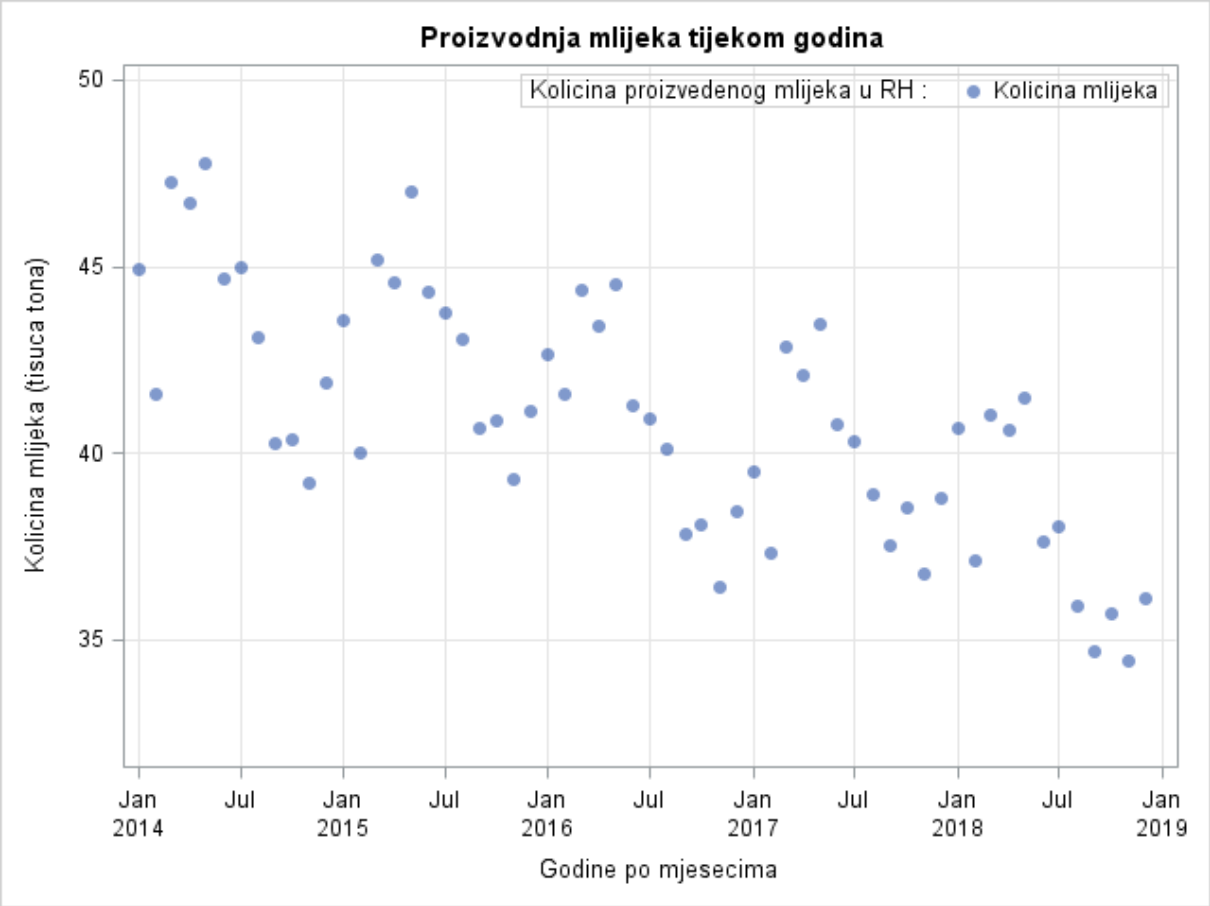
SCATTER X=nezavisna varijabla Y=zavisna varijabla </ opcije>;

RUN;

SAS primjer za raspršeni grafikon

```
TITLE 'Proizvodnja mlijeka tijekom godina';
PROC SGPLOT DATA=dipl.Eurostat;
SCATTER X=time Y=value / MARKERATTRS=(SYMBOL=CIRCLEFILLED) TRANSPARENCY=0.5;
XAXIS LABEL='Godine po mjesecima' GRID;
YAXIS MIN=32 MAX=50 LABEL='Kolicina mlijeka (tisuca tona)' GRID;
KEYLEGEND / LOCATION=INSIDE POSITION=TOPRIGHT ACROSS=1 TITLE='Kolicina proizvedenog mlijeka u RH :';
RUN;
```

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. U naredbi SCATTER koristimo vrijednosti koje nam predstavljaju vrijednosti nezavisne (*time*) i zavisne varijable (*value*). Unutar SERIES naredbe MARKERATTRS opcijom mijenjamo izgled simbola opcijom SYMBOL, odabrali smo prikaz ispunjenog kruga, tako da nam vrijednosti budu jasno prikazane. TRANSPARENCY opcijom određujemo razinu prozirnosti točaka koje predstavljaju opažanja na grafikonu poprima vrijednosti od 0 do 1 gdje je 1 potpuno prozirno, a 0 potpuno ispunjeno. U našem grafikonu odabrali smo 0.5 zbog čega je prikaz djelomično proziran. Unutar naredbi XAXIS i YAXIS opcijama LABEL, MIN, MAX, GRID određujemo način prikaza podataka. Opcija LABEL daje naziv varijabli. Opcijom MIN određujemo najnižu vrijednost na osi. Opcijom MAX određujemo najvišu vrijednost na osi. Opcijom GRID prikazujemo osi koordinatnog sustava. Unutar naredbe KEYLEGEND određujemo način prikazivanja legende koristeći opcije LOCATION, POSITION i ACROSS. LOCATION opcija koristi se za promjenu lokaciju legende, dakle nalazi li se legenda unutar ili van grafikona. Opcija POSITION govori nam o smještaju grafikona, smještaj može biti gore, dolje, desno, lijevo i na sredini. Opcijom ACROSS korigiramo prikaz koloni u legendi, ACROSS 1 prikazuje nam varijable u jednoj koloni. Opcijom TITLE dodali smo naziv legendi.



Slika 1. Prikaz raspršenog grafikona SGPLOT procedurom

Objašnjenje grafikona: Na slici 1 prikazan je raspršeni grafikon. Na ovom grafikonu prikazali smo količinu proizvedenog mlijeka u Republici Hrvatskoj po mjesecima između 2014. i 2019. godine. Na Y osi prikazana je količina proizvedenog mlijeka, dok su na X osi prikazane godine po mjesecima. U legendi je prikazana vrijednost varijable Količina proizvedenog mlijeka u RH koja je prikazana na grafikonu. Plavim točkama prikazani su podatci za količinu proizvedenog mlijeka u RH između 2014. i 2019. godine. Možemo vidjeti da je količina proizvedenog mlijeka svake godine najviša početkom godine te se smanjuje prema sredini godine i ponovno raste prema kraju godine.

3.1.2. Linijski grafikon (eng. Series plot)

Prema Delwiche i Slaughter (2012) linijski grafikon sličan je raspršenom grafikonu samo što se ne označava svako opažanje na grafikonu nego ih se povezuje linijom. Isti autori tvrde da linijski grafikon ima smisla koristiti kada se podaci trebaju prikazati nekim određenim redoslijedom. Prema Matange i Heath (2011) linija prolazi kroz svako opažanje te tvori krivulju koja nam govori o vrijednostima varijabli, spajanjem više linijskih grafikona u jedan možemo uspoređivati kretanje nekih podataka kroz neki vremenski period kod više jedinki ili grupa.

Delwiche i Slaughter (2011) tvrde da je osnovni oblik linijskog grafikona uz raspršeni grafikon glavni način za prikaz neobrađenih podataka.

Sintaksa SAS programa prema Matange i Heath (2012):

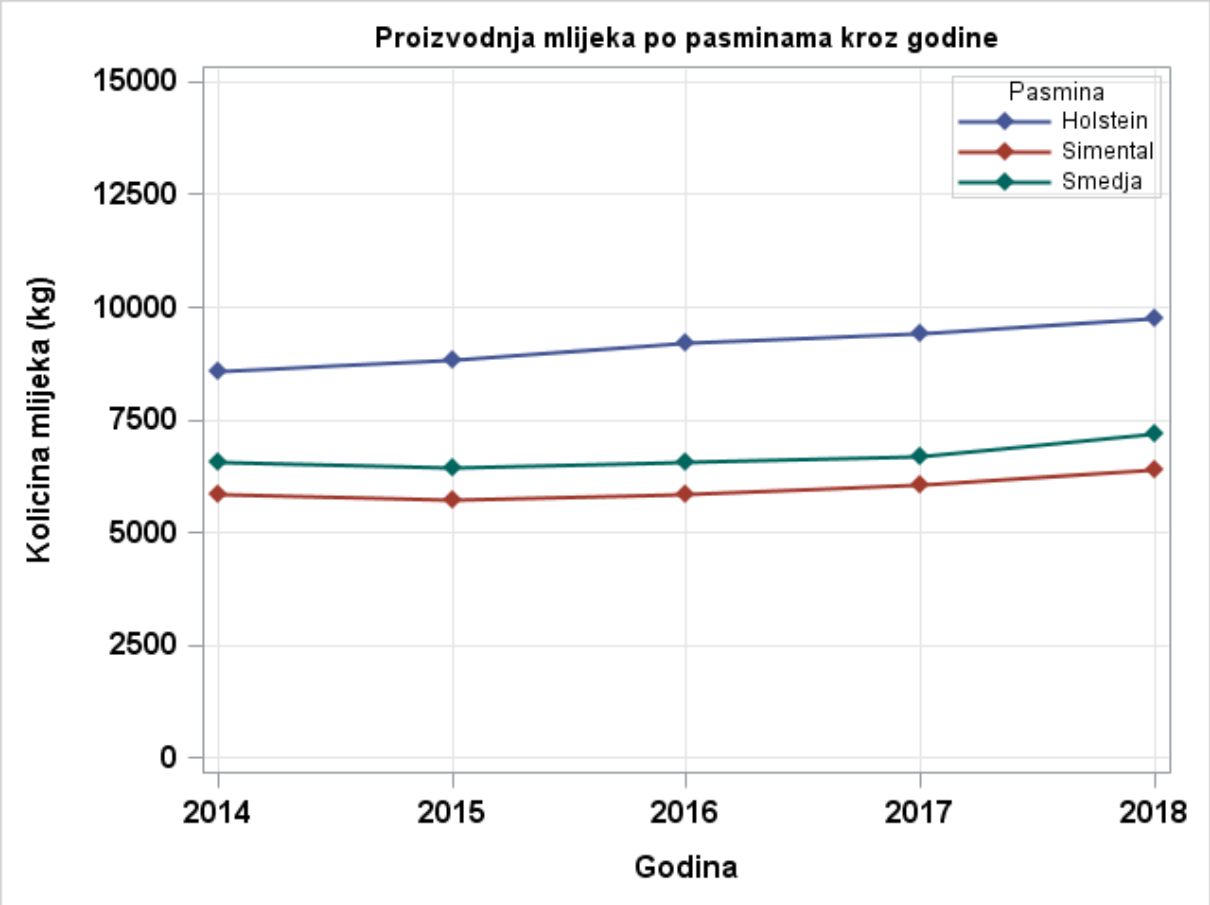
```
PROC SGPLOT;
SERIES X=column Y=column </ OPTIONS>;
RUN;
```

SAS primjer linijskog grafikona

```
TITLE 'Proizvodnja mlijeka po pasminama kroz godine';
PROC SGPLOT DATA=dipl.Hpa_podaci2;
SERIES X=godina Y=my /
GROUP=pasmina MARKERS MARKERATTRS=(SYMBOL=diamondfilled)
LINEATTRS=(THICKNESS=2);
KEYLEGEND / LOCATION=inside POSITION=topright ACROSS=1;
YAXIS MIN=0 MAX=15000
LABEL='Kolicina mlijeka (kg)' LABELATTRS=(WEIGHT=bold SIZE=12)
VALUEATTRS=(WEIGHT=bold SIZE=12) GRID;
XAXIS LABEL='Godina' LABELATTRS=(WEIGHT=bold SIZE=12)
VALUEATTRS=(WEIGHT=bold SIZE=12) GRID;
RUN;
```

Objašnjenje: TITLE naredbom dajemo naziv grafikonu. Korištena je SGPLOT procedura gdje su u naredbi SERIES označene vrijednosti nezavisne (*godina*) i zavisne varijable (*my*). Unutar SERIES naredbe GROUP opcijom grupiramo po kojoj će se varijabli (*pasmina*) prikazivati podaci prikazani na X i Y osi te opcijama MARKERS, MARKERATTRS, LINEATTRS utječemo na prikaz podataka u grafikonu. MARKERS opcijom definiramo uključujemo prikaz opažanja na grafikonu, jer je u postavkama da ih se ne prikazuje. MARKERATTRS opcijom mijenjamo izgled simbola opcijom SYMBOL, odabrali smo prikaz ispunjenog dijamanta, tako da nam vrijednosti budu jasno prikazane. LINEATTRS opcijom određujemo debljinu linije opcijom THICKNESS, odabrali smo veličinu 2 da nam linija bude dvostruko deblja od početnih postavki. Unutar naredbe KEYLEGEND određujemo način prikazivanja legende koristeći opcije LOCATION, POSITION i ACROSS. LOCATION opcija koristi se za promjenu lokaciju legende, dakle nalazi li se legenda unutar ili van grafikona. Opcija POSITION govori nam o smještaju grafikona, smještaj može biti gore, dolje,

desno, lijevo i na sredini. Opcijom ACROSS korigiramo prikaz koloni u legendi, ACROSS 1 prikazuje nam varijable u jednoj koloni. Unutar naredbi YAXIS i XAXIS koristeći MIN, MAX, LABEL, LABELATTRS, VALUEATTRS, GRID opcije određujemo način prikaza podataka. Opcija LABEL daje naziv varijabli. Opcijom MIN određujemo najnižu vrijednost na osi. Opcijom MAX određujemo najvišu vrijednost na osi. Opcijom GRID prikazujemo osi koordinatnog sustava. Opcijom LABELATTRS korigiramo debljinu vrijednosti varijable pomoću opcije WEIGHT, i veličinu slova opcijom SIZE. Opcijom VALUEATTRS korigiramo debljinu opažanja na grafikonu opcijom WEIGHT, dok opcijom SIZE mijenjamo veličinu opažanja.



Slika 2. Prikaz linijskog grafikona SGPLOT procedurom

Objašnjenje grafikona: Na slici 2 prikazan je linijski grafikon. Na ovom grafikonu prikazali smo proizvodnju mlijeka po pasminama u između 2014. i 2018. godine. Budući da smo podatke grupirali prema pasminama, plavom bojom je prikazana Holstein pasmina, crvenom bojom Simentalska, dok je zelenom bojom prikazana Smedja pasmina. Na Y osi prikazana je količina proizvedenog mlijeka izražena u kilogramima, dok su na X osi prikazane godine. Na legendi je prikazana klasifikacija pasmina prema bojama. Iz ovog grafikona možemo vidjeti da je prosječna količina proizvedenog mlijeka porasla kod svih pasmina. Kod

Holstein goveda je vidljiv najveći porast proizvodnje, dok je kod Simentalske i Smedje pasmine vidljiv blagi porast u proizvodnosti nakon 2016. godine.

3.1.3. Mjehuričasti grafikoni (eng. Bubble plot)

Prema Lafler-u (2018) mjehuričasti plotovi omogućuju nam trodimenzionalni način prikaza podataka gdje se točke mijenja mjehurićima. Lafler tvrdi da su nam za prikaz mjehurića potrebne koordinate točaka jer se na temelju tih podataka može prikazati opažanje na grafikonu. Glavna razlika u odnosu na raspršeni grafikon je treća varijabla koja nam govori o veličini ili radijusu mjehurića, odnosno koliko vrijednosti se nalazi unutar određenog razreda (Matange i Heath 2011).

Sintaksa SAS programa prema Matange i Heath (2011)

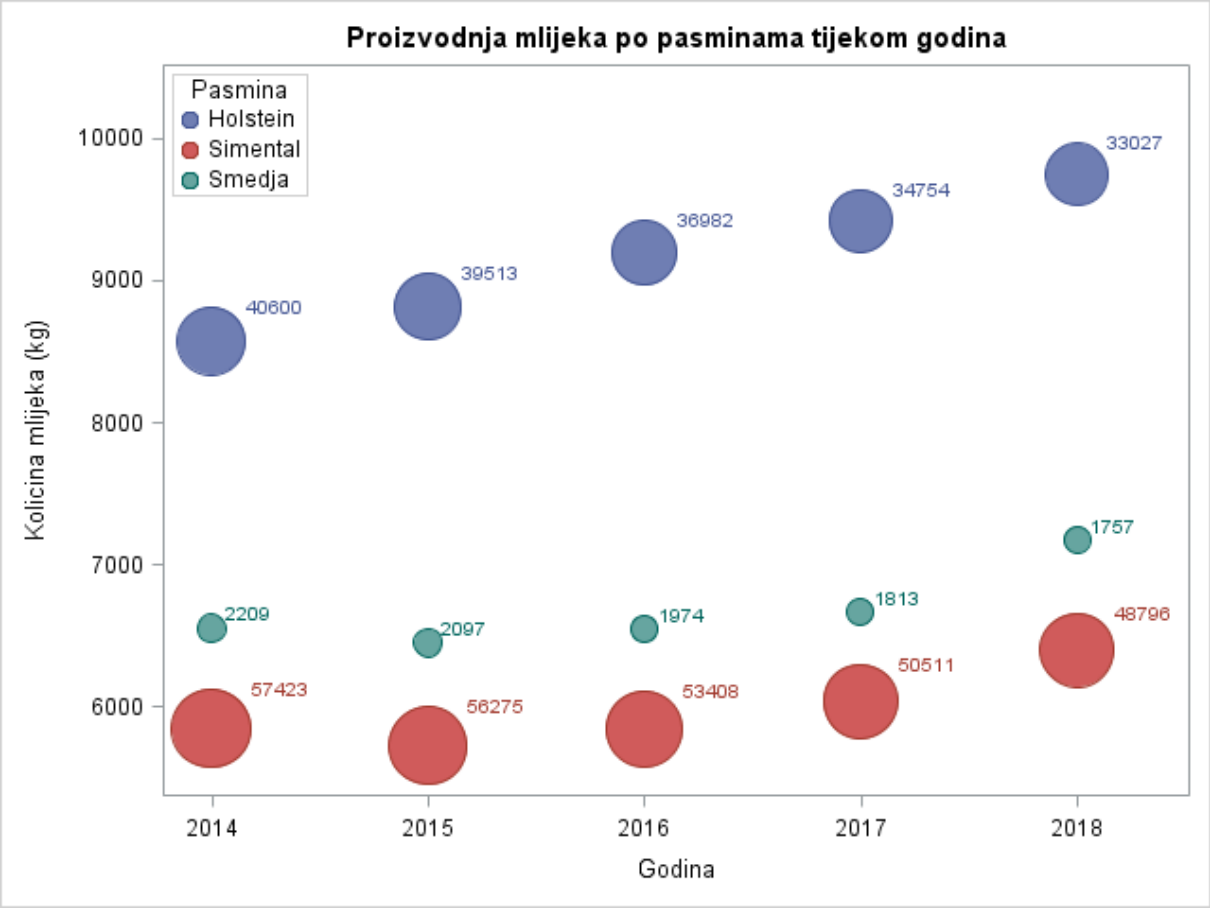
```
PROC SGPLOT;
BUBBLE X=nezavisna varijabla Y=zavisna varijabla SIZE=varijabla veličine mjehurića
</opcije>;
RUN;
```

SAS primjer mjehuričastog grafikona

```
TITLE 'Proizvodnja mlijeka tijekom godina';
PROC SGPLOT DATA=dipl.Hpa_podaci2;
BUBBLE X=godina Y=my SIZE=stado / GROUP=pasmina DATALABEL=stado;
XAXIS LABEL='Godina';
YAXIS LABEL='Kolicina mlijeka';
KEYLEGEND / LOCATION=inside POSITION=topleft ACROSS=1;
RUN;
```

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. U BUBBLE naredbi koristimo vrijednosti koje nam predstavljaju nezavisnu varijablu (*godina*), zavisnu varijablu (*my*), i varijablu koja nam govori o veličini mjehurića (*stado*). Unutar BUBBLE naredbe opcijom GROUP podatke dijelimo u tri grupe pomoću varijable (*pasmina*), dok opcijom DATALABEL određujemo varijablu (*stado*) čija će vrijednost biti prikazana kraj mjehurića. Unutar naredbi XAXIS i YAXIS opcijom LABEL definiramo nazive varijabli koje će se nalaziti na X i Y osima. Unutar naredbe KEYLEGEND određujemo način na koji će nam se prikazivati legenda te gdje će se nalaziti. LOCATION opcija koristi se za promjenu lokaciju

legende, dakle nalazi li se legenda unutar ili van grafikona. Opcija POSITION govori nam o smještaju grafikona, smještaj može biti desno, lijevo i na sredini. Opcijom ACROSS korigiramo prikaz koloni u legendi, ACROSS 1 prikazuje nam varijable u jednoj koloni.



Slika 3. Prikaz mjehuričastog grafikona SGPLOT procedurom

Objašnjenje grafikona: Na slici 3 prikazan je mjehuričasti grafikon. Na ovom grafikonu prikazali smo proizvodnju mlijeka po pasminama između 2014. i 2018. godine. Budući da smo podatke grupirali prema pasminama, plavom bojom je prikazana Holstein pasmina, crvenom bojom Simetnalska, dok je zelenom bojom prikazana Smedja pasmina. Na Y osi prikazana je količina proizvedenog mlijeka izražena u kilogramima, dok su na X osi prikazane godine. Na legendi je prikazana klasifikacija pasmina prema bojama. Podaci ispisani kraj mjehura govore nam o broju grla u kontroli mliječnosti, što nam različite veličine mjehura slikovno prikazuju svojom veličinom. Iz ovog grafikona možemo vidjeti da je prosječna količina proizvedenog mlijeka porasla kod svih pasmina. Kod Holstein goveda je vidljiv najveći porast proizvodnje, dok je kod Simentalske i Smedjih pasmina vidljiv blagi porast u proizvodnosti nakon 2016. godine.

3.2. Grafikoni prikaza modela i intervala pouzdanosti

Prema Matange i Heath (2011) grafikoni prikaza modela i intervala pouzdanosti koriste se za prikaz regresijskih modela, loess krivulja, penalized B-spline krivulja, te elipsa.

SAS program sintaksa za grafikone prikaza modela i intervala pouzdanosti**:**

PROC SGPLOT;

NAZIV-NAREDBE X=nezavisna varijabla Y=zavisna varijabla </ opcije>;

RUN;

Ovisno o tome koju vrstu krivulje želimo prikazati naziv naredbe može biti: REG, LOESS i PBSPLINE.

3.2.1. Regresijski grafikon

Prema Matange i Heath (2011) regresijski grafikon koristi se za prikaz linearne regresije te pokazuje linearni i nelinearni prikaz podataka. Tako da se na temelju opažanja u koordinatnom sustavu definira regresijski pravac iz modela.

Sintaksa SAS programa prema Matange i Heath (2011)

PROC SGPLOT;

REG X=nezavisna varijabla Y=zavisna varijabla </ opcije>;

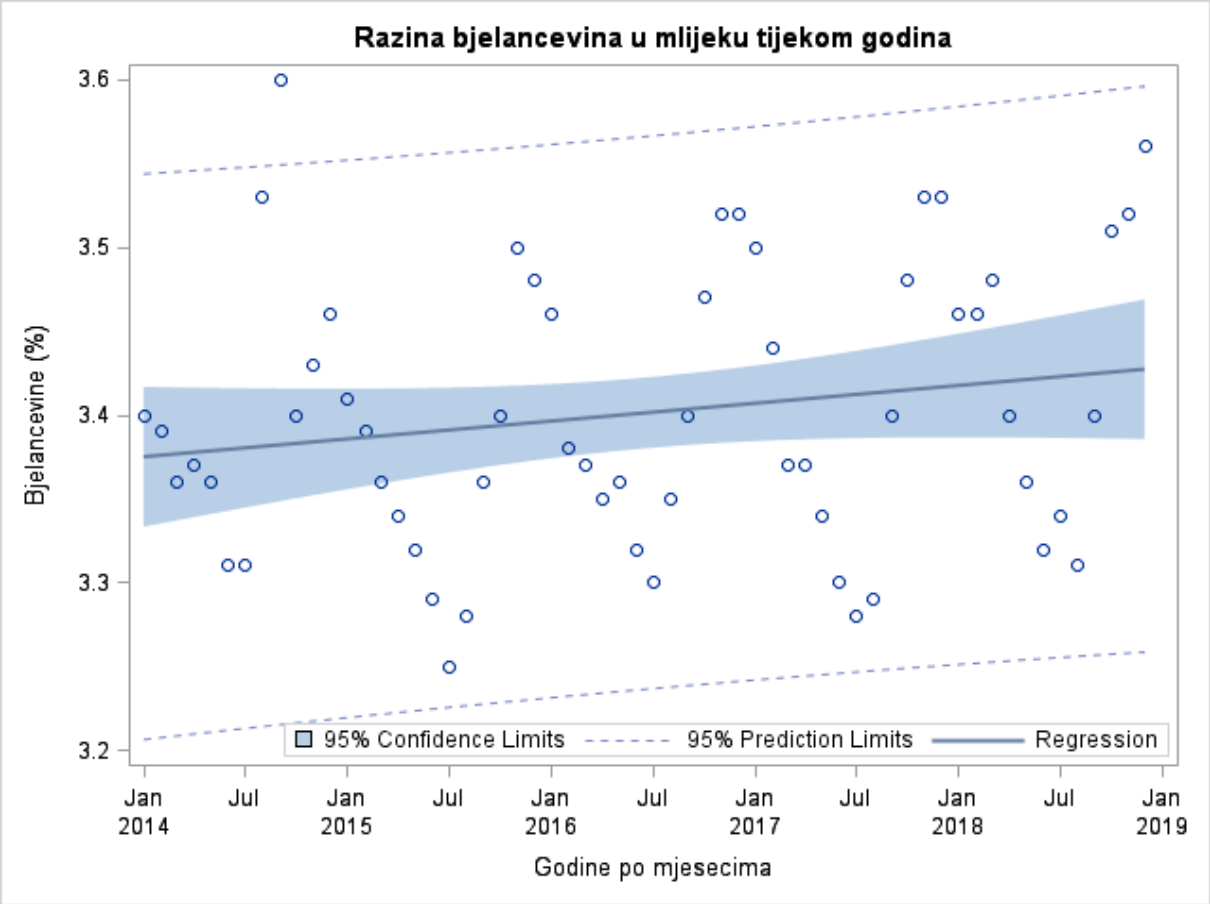
RUN;

SAS primjer regresijskog grafikona za bjelančevine

```
TITLE 'Razina bjelancevina u mlijeku tijekom godina';
PROC SGPLOT DATA=dipl.Eurostat;
REG X=TIME Y=prot / CLI CLM;
KEYLEGEND /LOCATION=inside;
XAXIS LABEL='Godine po mjesecima';
YAXIS LABEL='Bjelancevine (%)';
RUN;
```

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. U naredbi REG koristimo vrijednosti koje nam predstavljaju nezavisnu (*TIME*) i zavisnu varijablu (*prot*). Opcijom CLI dodajemo intervale pouzdanosti za svaku predviđenu vrijednost, dok CLM opcijom dodajemo intervale pouzdanosti za prosječne predviđene vrijednosti. Unutar naredbe KEYLEGEND

određujemo način na koji će nam se prikazivati legenda te gdje će se nalaziti pomoću LOCATION opcije. LOCATION opcija koristi se za promjenu lokaciju legende, dakle nalazi li se legenda unutar ili van grafikona. Unutar naredbi XAXIS i YAXIS opcijom LABEL definiramo nazive varijabli koje će se nalaziti na X i Y osima.



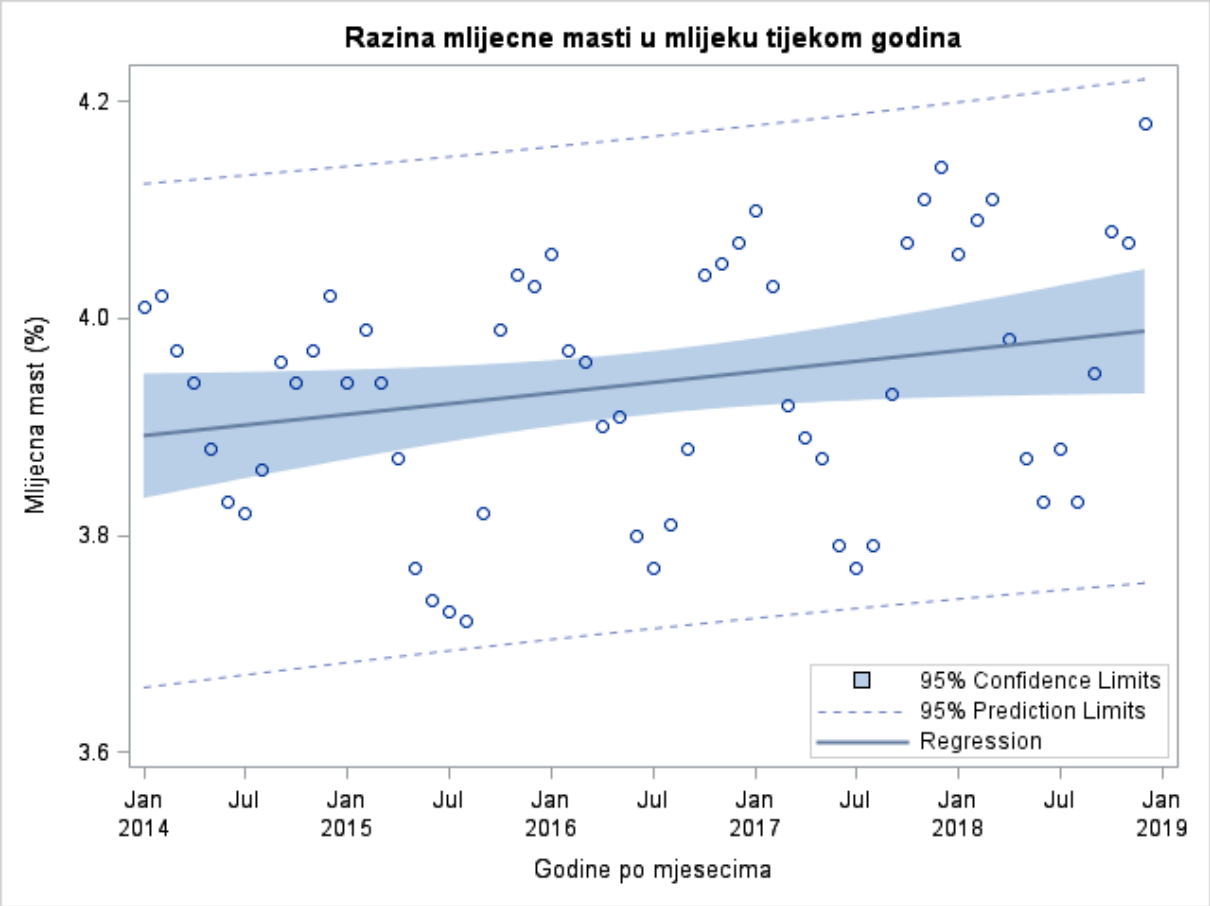
Slika 4. Prikaz regresijskog grafikona SGPLOT procedurom

Objašnjenje grafikona: Na slici 4 prikazan je regresijski grafikon. Na ovom grafikonu prikazali smo razinu bjelančevina u mlijeku u između 2014. i 2019. godine. Opažanja vidljiva na grafikonu prikazuju nam razinu bjelančevina u mlijeku izraženu u postotcima. Plava linija pokazuje nam regresijski pravac. Isprekidane linije prikazuju nam intervale pouzdanosti za svaku predviđenu vrijednost. Dio grafikona ispunjen plavom bojom prikazuje nam intervale pouzdanosti za prosječne predviđene vrijednosti. Na Y osi prikazana je razina bjelančevina u mlijeku izražena u postotcima, dok su na X osi prikazane godine. Legenda je prikazana vodoravno budući da nije korištena ACROSS opcija. U legendi su označeni prikazi intervala pouzdanosti za svaku predviđenu vrijednost koji iznosi 95 %, intervala pouzdanosti za prosječne predviđene vrijednosti koji iznosi 95 %, regresijskog pravca. Iz ovog grafikona možemo vidjeti da se razina bjelančevina u mlijeku gotovo neznatno povećala.

SAS primjer regresijskog grafikona za mliječnu mast

```
TITLE 'Razina mlijecne masti u mlijeku tijekom godina';
PROC SGPLOT DATA=dipl.Eurostat;
REG X=TIME Y=mm / CLI CLM;
KEYLEGEND /LOCATION=inside ACROSS=1;
XAXIS LABEL='Godine po mjesecima';
YAXIS LABEL='Mlijecna mast (%)';
RUN;
```

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. U naredbi REG koristimo vrijednosti koje nam predstavljaju nezavisnu (*TIME*) i zavisnu varijablu (*mm*). Opcijom CLI dodajemo intervale pouzdanosti za svaku predviđenu vrijednost, dok CLM opcijom dodajemo intervale pouzdanosti za prosječne predviđene vrijednosti. Unutar naredbe KEYLEGEND određujemo način na koji će nam se prikazivati legenda te gdje će se nalaziti pomoću LOCATION opcije. LOCATION opcija koristi se za promjenu lokaciju legende, dakle nalazi li se legenda unutar ili van grafikona. Opcijom ACROSS korigiramo prikaz koloni u legendi, ACROSS 1 prikazuje nam varijable u jednoj koloni. Unutar naredbi XAXIS i YAXIS opcijom LABEL definiramo nazive varijabli koje će se nalaziti na X i Y osima.



Slika 5. Prikaz regresijskog grafikona SGPLOT procedure

Objašnjenje grafikona: Na slici 5 prikazan je regresijski grafikon. Na ovom grafikonu prikazali smo razinu mliječne masti u mlijeku u između 2014. i 2019. godine. Opažanja vidljiva na grafikonu prikazuju nam razinu mliječne masti u mlijeku izraženu u postotcima. Plava linija pokazuje nam regresijski pravac. Isprekidane linije prikazuju nam intervale pouzdanosti za svaku predviđenu vrijednost. Dio grafikona ispunjen plavom bojom prikazuje nam intervale pouzdanosti za prosječne predviđene vrijednosti. Na Y osi prikazana je razina mliječne masti u mlijeku izražena u postotcima, dok su na X osi prikazane godine po mjesecima. U legendi su označeni prikazi intervala pouzdanosti za svaku predviđenu vrijednost koji iznosi 95 %, intervala pouzdanosti za prosječne predviđene vrijednosti koji iznosi 95 %, regresijskog pravca. Iz ovog grafikona možemo vidjeti da se razina mliječne masti u mlijeku gotovo neznatno povećala.

3.2.2. Grafikon loess funkcije

Prema Matange i Heath (2011) Loess funkcija nam prikazuje prilagođenu krivulju na temelju rasporeda točaka na grafikonu.

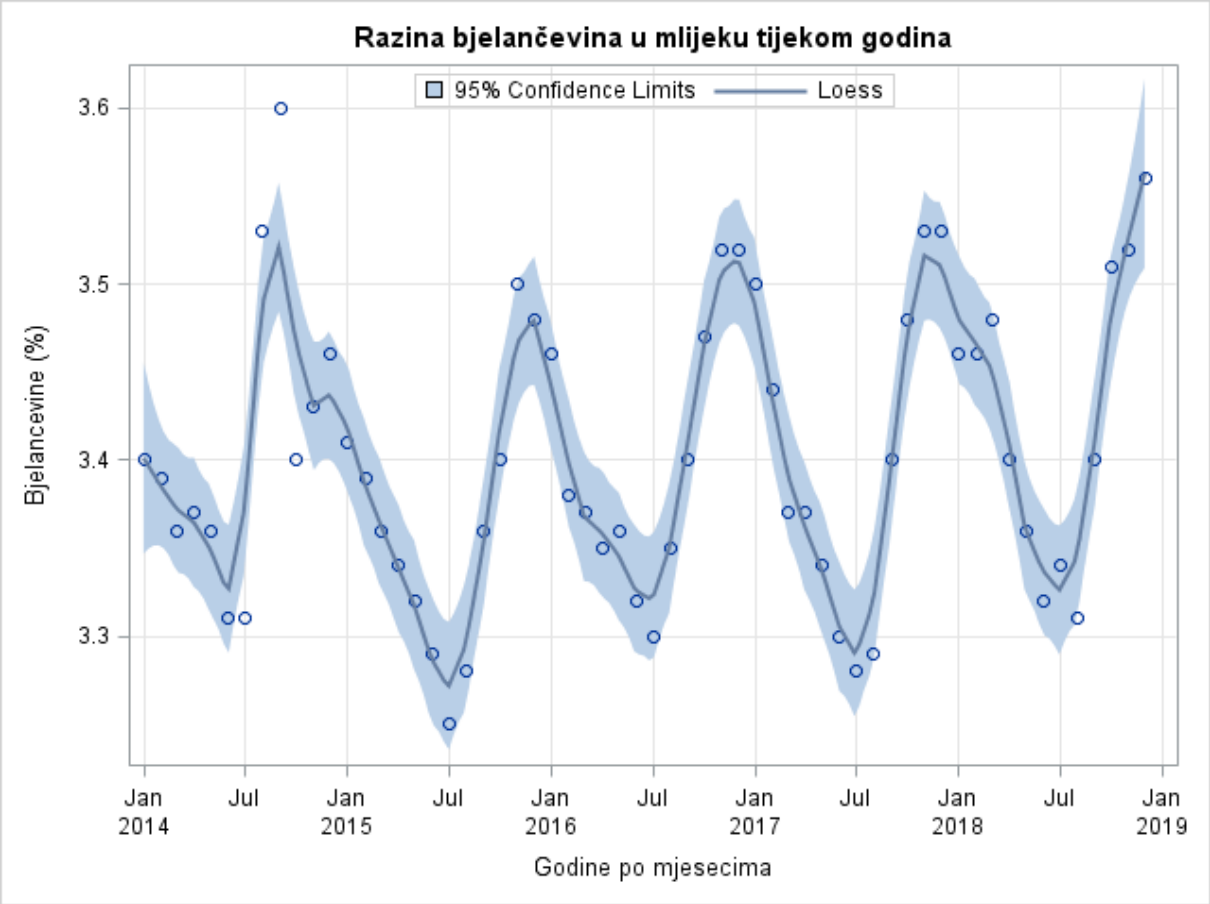
Sintaksa SAS programa prema Matange i Heath (2011)

```
PROC SGPLOT;
LOESS X=nezavisna varijabla Y=zavisna varijabla </ opcije>;
RUN;
```

SAS primjer grafikona LOESS funkcije za bjelančevine

```
TITLE 'Razina bjelančevina u mlijeku tijekom godina';
PROC SGPLOT DATA=dipl.Eurostat;
LOESS X=TIME Y=prot / CLM;
XAXIS LABEL='Godine po mjesecima' GRID;
YAXIS LABEL='Bjelancevine (%)' GRID;
KEYLEGEND / LOCATION=inside POSITION=top;
RUN;
```

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. U LOESS naredbi koristimo vrijednosti koje nam predstavljaju nezavisnu (*TIME*) i zavisnu varijablu (*prot*). Opcijom CLM dodajemo intervale pouzdanosti za prosječne predviđene vrijednosti. Unutar naredbi XAXIS i YAXIS opcijom LABEL definiramo nazive varijabli koje će se nalaziti na X i Y osima, dok GRID opcijom dodajemo linije koordinatnog sustava. Unutar naredbe KEYLEGEND određujemo način na koji će nam se prikazivati legenda te gdje će se nalaziti. LOCATION opcija koristi se za promjenu lokaciju legende, dakle nalazi li se legenda unutar ili van grafikona. Opcija POSITION govori nam o poziciji legende, legenda može biti smještena gore, dolje, desno, lijevo i na sredini.



Slika 6. Prikaz Loess funkcije SGPLOT procedurom

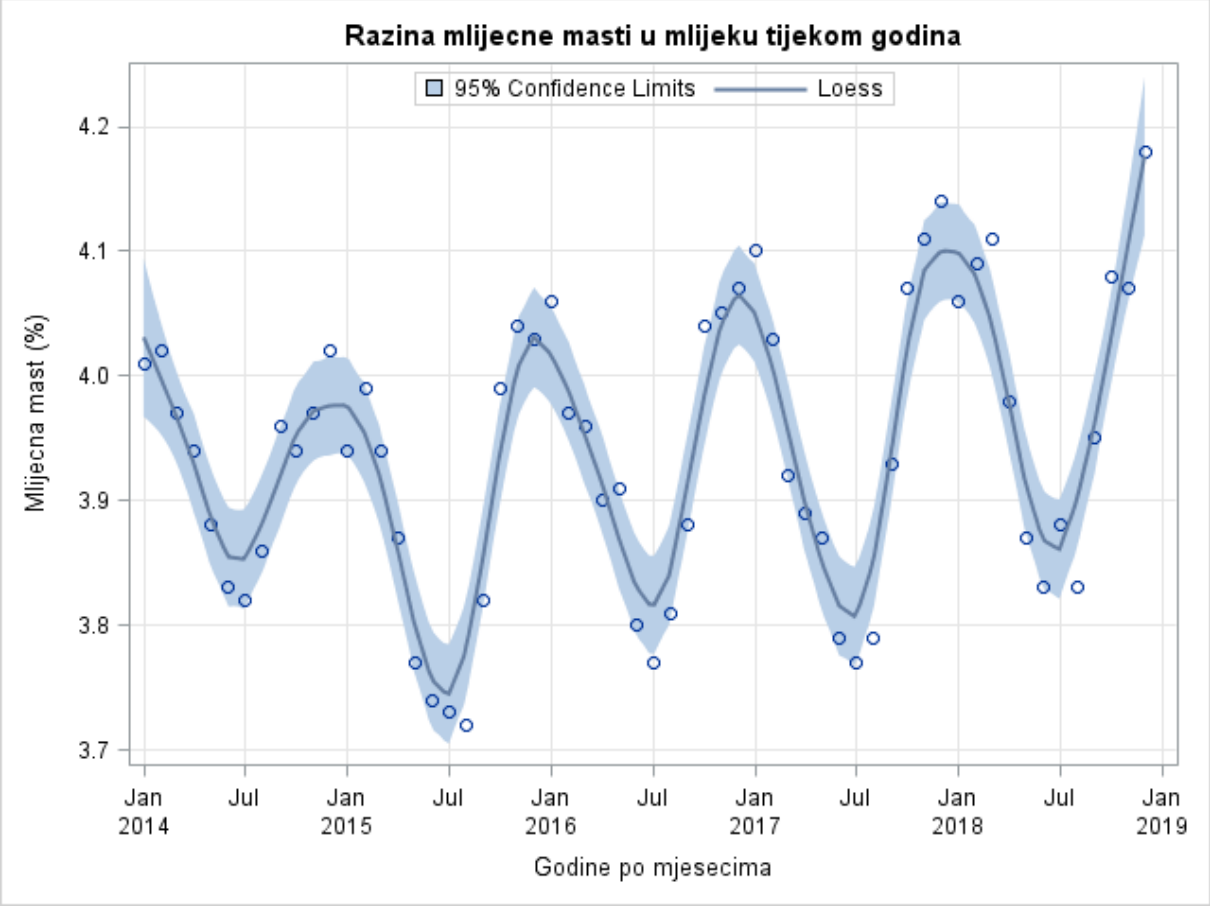
Objašnjenje grafikona: Na slici 6 prikazana je Loess funkcija. Na ovom grafikonu prikazali smo razinu bjelančevina u mlijeku između 2014. i 2019. godine. Opažanja vidljiva na grafikonu prikazuju nam razinu bjelančevina u mlijeku izraženu u postotcima. Plava linija pokazuje nam regresijski pravac. Dio grafikona ispunjen plavom bojom prikazuje nam intervale pouzdanosti za prosječne predviđene vrijednosti. Na Y osi prikazana je razina bjelančevina u mlijeku izražena u postotcima, dok su na X osi prikazane godine po mjesecima. Legenda je prikazana vodoravno budući da nije korištena ACROSS opcija. Iz ovog grafikona možemo vidjeti da su se najviše vrijednosti bjelančevina u mlijeku gotovo neznatno povećale tijekom 5 godina. Isto tako možemo vidjeti da je razina bjelančevina najviša početkom godine te da se smanjuje do 7. mjeseca, kada je i najniža u godini, nakon čega do kraja godine raste.

SAS primjer grafikona LOESS funkcije za mliječnu mast

```
TITLE 'Razina mlijecne masti u mlijeku tijekom godina';
PROC SGPLOT DATA=dipl.Eurostat;
LOESS X=TIME Y=mm / CLM;
XAXIS LABEL='Godine po mjesecima' GRID;
```

YAXIS LABEL='Mlijecna mast (%)' GRID;
KEYLEGEND / LOCATION=inside POSITION=top;
RUN;

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. U LOESS naredbi koristimo vrijednosti koje nam predstavljaju nezavisnu (*TIME*) i zavisnu varijablu (*mm*). Opcijom CLM dodajemo intervale pouzdanosti za prosječne predviđene vrijednosti. Unutar naredbi XAXIS i YAXIS opcijom LABEL definiramo nazive varijabli koje će se nalaziti na X i Y osima, dok GRID opcijom dodajemo linije koordinatnog sustava. Unutar naredbe KEYLEGEND određujemo način na koji će nam se prikazivati legenda te gdje će se nalaziti. LOCATION opcija koristi se za promjenu lokaciju legende, dakle nalazi li se legenda unutar ili van grafikona. Opcija POSITION govori nam o poziciji legende, legenda može biti smještena gore, dolje, desno, lijevo i na sredini.



Slika 7. Prikaz Loess krivulje SGPLOT procedurom

Objašnjenje grafikona: Na slici 7 prikazana je Loess funkcija. Na ovom grafikonu prikazali smo razinu mliječne masti u mlijeku između 2014. i 2019. godine. Opažanja vidljiva na grafikonu prikazuju nam razinu mliječne masti u mlijeku izraženu u postotcima. Plava linija pokazuje nam regresijski pravac. Dio grafikona ispunjen plavom bojom prikazuje nam intervale pouzdanosti za prosječne predviđene vrijednosti. Na Y osi prikazana je razina mliječne masti u mlijeku izražena u postotcima, dok su na X osi prikazane godine po mjesecima. Legenda je prikazana vodoravno budući da nije korištena ACROSS opcija. Iz ovog grafikona možemo vidjeti da su se najviše vrijednosti mliječne masti u mlijeku gotovo neznatno povećale tijekom 5 godina. Isto tako možemo vidjeti da je razina mliječne masti najviša početkom godine te da se smanjuje do 7. mjeseca, kada je i najniža u godini, nakon čega do kraja godine raste.

3.3. Distribucijski grafikoni

3.3.1. Histogrami i krivulja funkcija gustoće

Prema Delwiche i Slaughter (2012) histogram se koristi za prikaz distribucije kontinuiranih podataka koji su podijeljeni u razrede, svaki razred ima interval te su na temelju toga opažanja uvrštena u određeni razred. Nadalje prema Delwiche i Slaughter, razlikuje se od stupčastih grafikona jer nema razmaka između stupaca.

Sintaksa SAS programa prema Matange i Heath (2011)

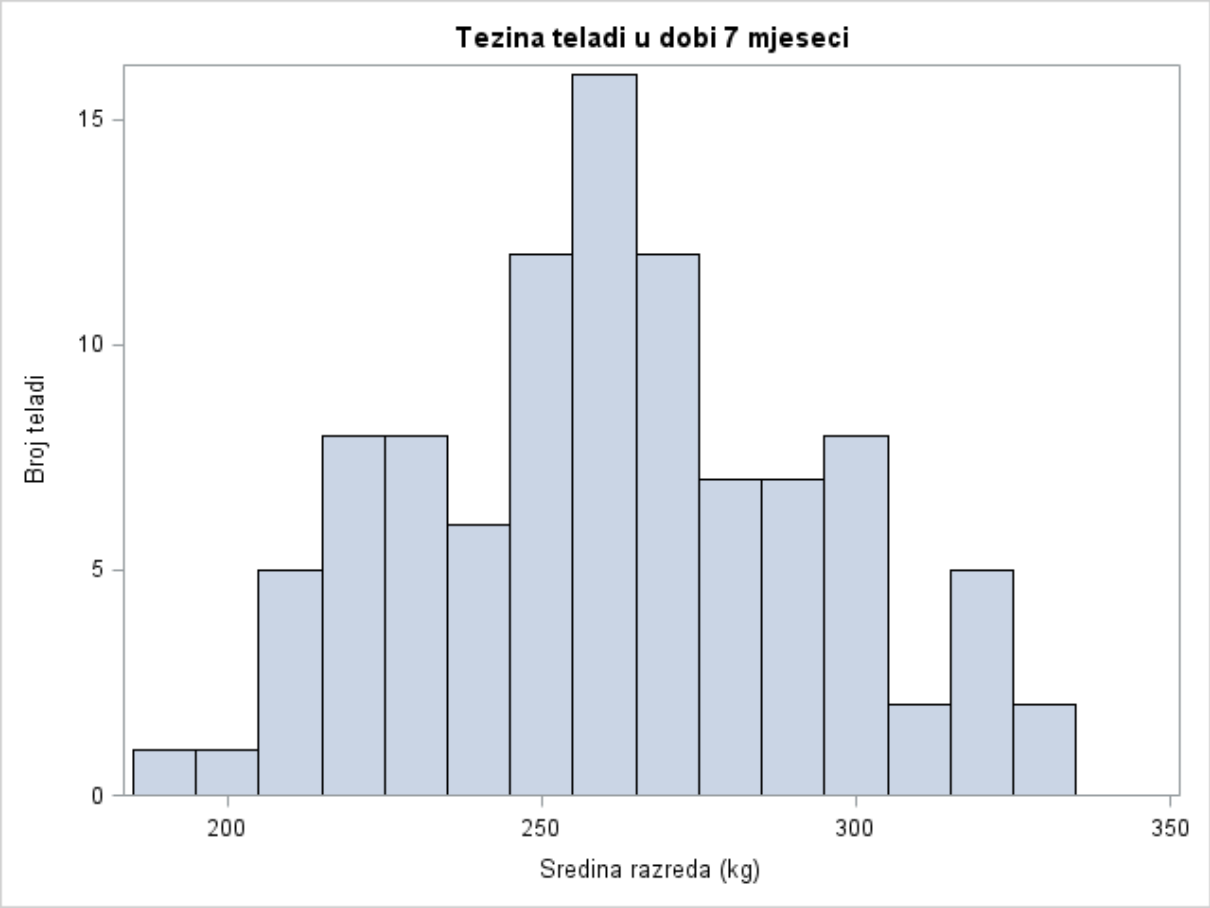
```
PROC SGPLOT;
HISTOGRAM naziv-varijable </ opcije>;
RUN;
```

SAS primjer histogram grafikona

```
TITLE 'Tezina teladi u dobi 7 mjeseci';
PROC SGPLOT DATA=dipl.hispodaci;
HISTOGRAM razred / NBINS=15 ;
XAXIS LABEL='Sredina razreda (kg)';
YAXIS LABEL='Broj teladi';
RUN;
```

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. Korištena je HISTOGRAM naredba. Za prikaz podataka koristimo kategoričku varijablu (*razred*). Unutar HISTOGRAM naredbe opcijom NBINS definiramo u koliko razreda želimo da su nam podaci prikazani. Naši

podaci prikazani su u 15 razreda. Unutar naredbi XAXIS i YAXIS opcijom LABEL definiramo nazive varijabli koje će se nalaziti na X i Y osima.



Slika 8. Prikaz Histograma SGPLOT procedurom

Objašnjenje grafikona: Na slici 8 prikazan je histogram. Na ovom histogramu prikazali smo težinu teladi u dobi od 7 mjeseci. Na Y osi prikazana je količina opažanja, dok je na X osi prikazana srednja vrijednost razreda izražena u kilogramima. Vrijednosti težine teladi prikazane su u 15 razreda. Iz ovog histograma možemo vidjeti da je težina teladi u dobi od 7 mjeseci najčešće oko 260 kilograma, dok se ekstremno teške i ekstremno lake teladi rjeđe pojavljuju.

3.3.2. Krivulja gustoće funkcije (eng. Density function)

Prema Delwiche i Slaughter (2012) krivuljom gustoće funkcije mogu se prikazati isti podaci koji se koriste za histogram. Krivulja gustoće nam prikazuje vjerojatnost, te vrijednost površine ispod krivulje iznosi 100 posto.

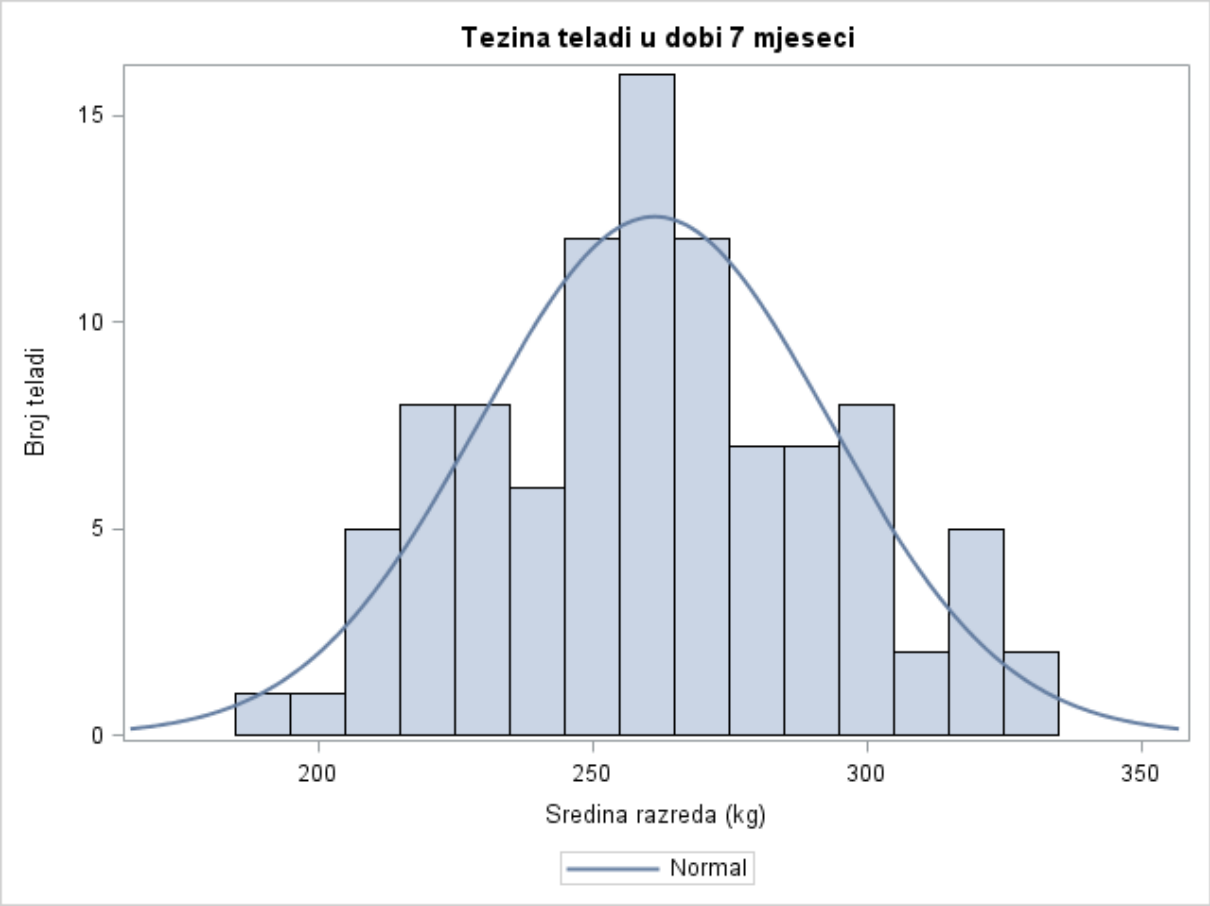
Sintaksa SAS programa prema Matange i Heath (2011)

```
PROC SGPLOT;
DENSITY naziv-varijable </ opcije>;
RUN;
```

SAS primjer grafikona krivulje funkcije gustoće

```
PROC SGPLOT DATA=dipl.hispodaci;
HISTOGRAM razred / NBINS=15 ;
DENSITY razred;
XAXIS LABEL='Sredina razreda (kg)';
YAXIS LABEL='Broj teladi';
RUN;
```

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. Korištena je HISTOGRAM naredba. Za prikaz podataka koristimo kategoričku varijablu (*razred*). Unutar HISTOGRAM naredbe opcijom NBINS definiramo u koliko razreda želimo da su nam podaci prikazani. DENSITY naredbom označavamo varijablu čije će nam vrijednosti krivulja gustoće prikazivati. Unutar naredbi XAXIS i YAXIS opcijom LABEL definiramo nazive varijabli koje će se nalaziti na X i Y osima.



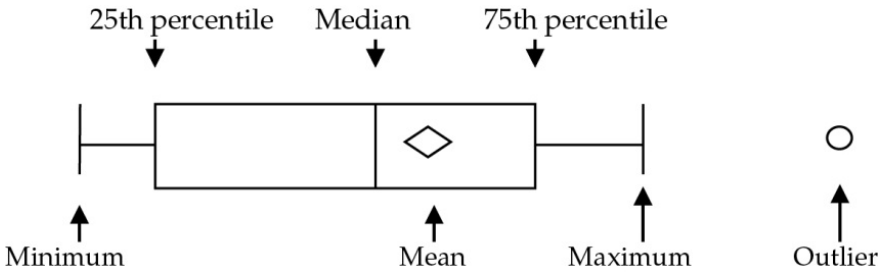
Slika 9. Prikaz Krivulje gustoće SGPLOT procedurom

Objašnjenje grafikona: Na slici 9 prikazan je histogram preko kojeg je prikazana krivulja gustoće. Na ovom histogramu prikazali smo težinu teladi u dobi od 7 mjeseci. Na Y osi prikazana je količina opažanja, dok je na X osi prikazana srednja vrijednost razreda izražena u kilogramima. Vrijednosti težine teladi prikazane su u 15 razreda. Na legendi je prikazan način prikaza krivulje gustoće. Iz ovog histograma možemo vidjeti da je težina teladi u dobi od 7 mjeseci najčešće oko 260 kilograma, dok se ekstremno teške i ekstremno lake teladi rjeđe pojavljuju. Krivulja gustoće nam pokazuje koje se vrijednosti najčešće pojavljuju. Budući da su nam podaci pravilno raspoređeni po razredima, krivulja gustoće izgleda poput normalne raspodjele.

Prema Delwiche i Slaughter (2012) HISTOGRAM i DENSITY naredbe mogu se koristiti zajedno, ali ne s drugim vrstama grafikona. Nadalje prema Delwich i Slaughter, moramo paziti na redoslijed pisanja koda, jer se preklapanja grafikona događaju prema redoslijedu napisanog koda. Tako da će druga upisana naredba biti prikazana na prvoj upisanoj naredbi na grafikonu.

3.3.3. Pravokutni grafikoni (eng. Box and Whishers plot)

Delwiche i Slaughter (2012) tvrde da poput histograma, pravokutni grafikoni prikazuju distribuciju kontinuiranih podataka. Svaki dio pravokutnog grafikona ima svoje značenje u razumijevanju podataka prikazanih u grafikonu. Krajevi pravokutnika označavaju gdje završava 25 i 75 posto podataka, linija unutar pravokutnika označava medijan odnosno gdje se nalazi 50 posto opažanja kada su opažanja poredana po veličini, a oznaka unutar pravokutnika označava aritmetičku srednju vrijednost, svaka vrijednost izvan minimalne i maksimalne vrijednosti pravokutnika je outlier tj. ekstremno opažanje (Delwiche i Slaughter, 2012). Zahvaljujući velikoj količini ugrađenih opcija koje su nam dostupne u SGPLOT proceduri, možemo napraviti okomite i vodoravne pravokutne grafikone.



Slika 10. Dijelovi pravokutnog grafikona

Izvor: Delwiche i Slaughter (2012)

Sintaksa SAS programa prema Matange i Heath (2011)

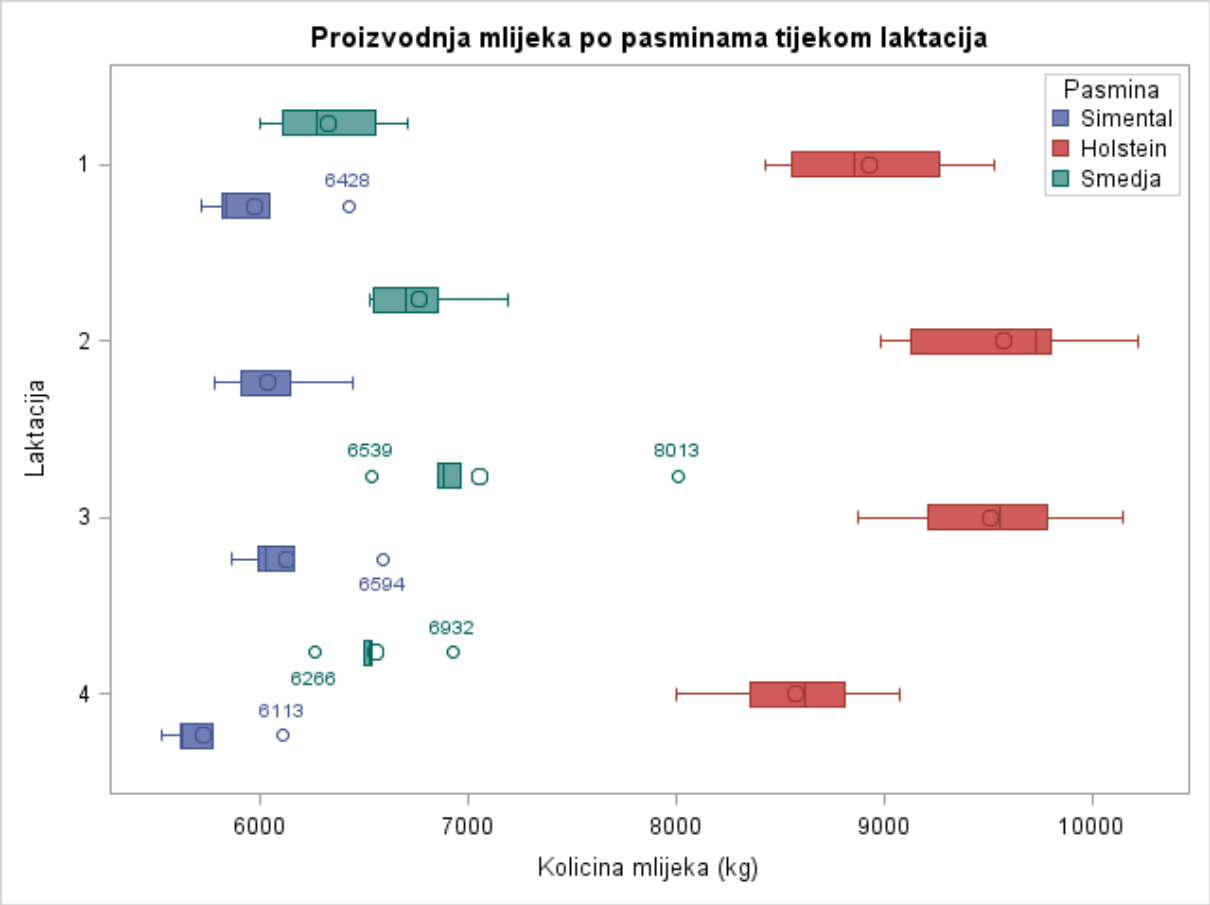
```
PROC SGPLOT;
VBOX/HBOX variable-name  </ OPTIONS>;
RUN;
```

SAS primjer vodoravnog pravokutnog grafikona

```
TITLE 'Proizvodnja mlijeka po pasminama tijekom laktacija';
PROC SGPLOT DATA=dipl.Hpa_podaci;
HBOX my /CATEGORY=laktacija GROUP=pasmina DATALABEL;
KEYLEGEND / LOCATION=INSIDE POSITION=TOPRIGHT ACROSS=1;
XAXIS LABEL='Kolicina mlijeka (kg)';
RUN;
```

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. Korištena je HBOX naredba čime radimo vodoravni pravokutni grafikon. Za prikaz podataka korištena je numerička varijabla (*my*). Opcija CATEGORY omogućuje nam prikaz podataka po kategorijama,

kategorizirali smo pomoću varijable (*laktacija*), tako da dobijemo pravokutni prikaz za svaku laktaciju. GROUP opcijom grupiramo prikaz podataka, grupirali smo pomoću varijable (*pasmina*), tako da nam se prikažu pravokutnici prema pasminama. Opcijom DATALABEL označavamo ispis vrijednosti koje se nalaze van pravukutnog grafikona, odnosno označavamo ekstremna opažanja. Unutar naredbe KEYLEGEND određujemo način na koji će nam se prikazivati legenda te gdje će se nalaziti. LOCATION opcija koristi se za promjenu lokaciju legende, dakle nalazi li se legenda unutar ili van grafikona. Opcija POSITION govori nam o smještaju grafikona, smještaj može biti desno, lijevo i na sredini. Opcijom ACROSS korigiramo prikaz koloni u legendi, ACROSS=1 prikazuje nam varijable u jednoj koloni.



Slika 11. Prikaz vodoravnog pravokutnog grafikona SGPLOT procedurom

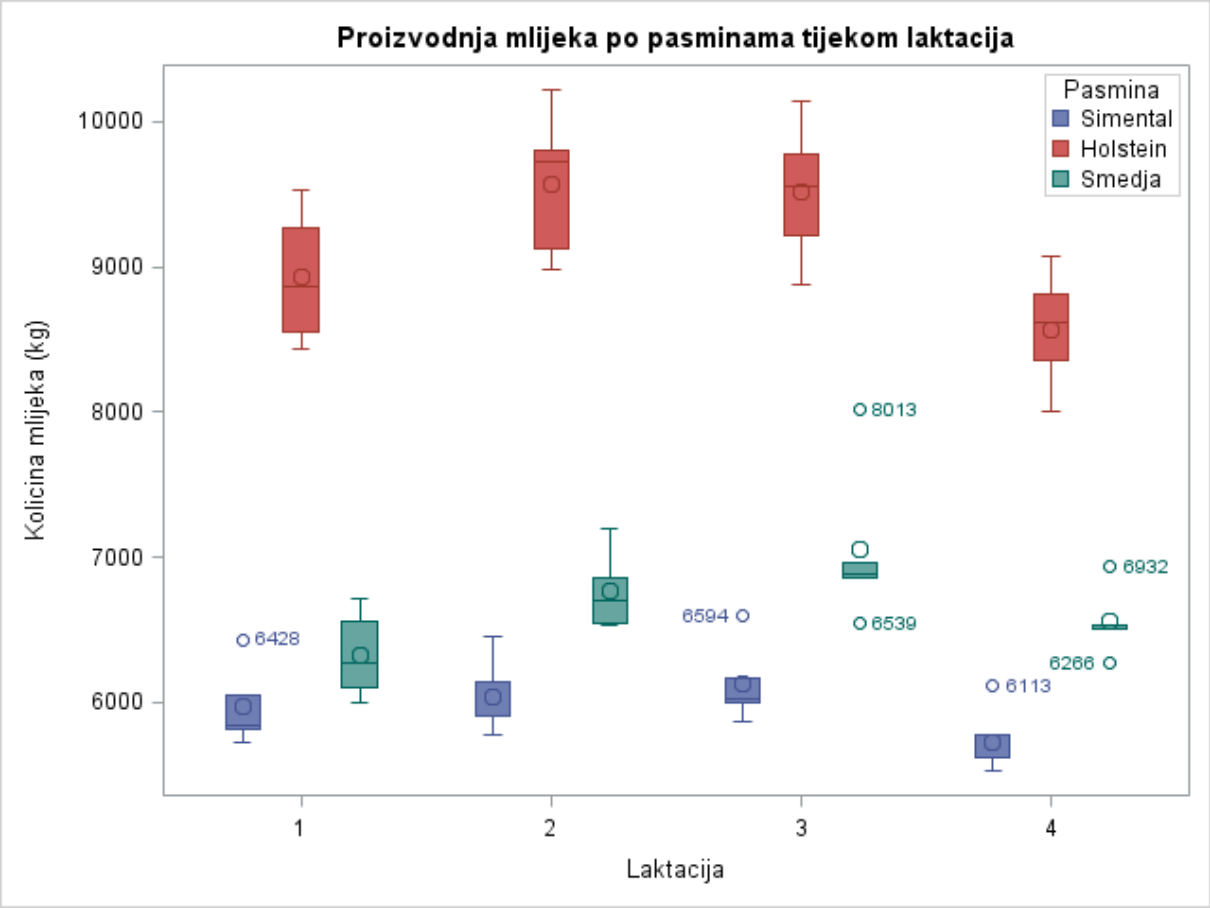
Objašnjenje grafikona: Na slici 11 prikazan je vodoravni pravokutni grafikon. Na ovom grafikonu prikazali smo prosječnu proizvodnju mlijeka po pasminama tijekom laktacija. Na Y osi prikazane su laktacije od 1 do 4, dok su na X osi prikazana količine proizvedenog mlijeka izražene u kilogramima. Na legendi je prikazana klasifikacija pasmina prema bojama. Budući da smo podatke grupirali prema pasminama, plavom bojom prikazana je Simentalska pasmina, crvenom bojom Holstein pasmina, dok su zelenom bojom prikazane Smedje pasmine. Zahvaljujući kategorizaciji podaci su nam prikazani prema laktacijama od 1 do 4 ovisno o

laktaciji u kojoj su se krave nalazile. Iz ovog grafikona možemo vidjeti da je najviša proizvodnja mlijeka u 2. i 3. laktaciji, dok je najniža proizvodnja u 4. laktaciji, te da Holstein goveda proizvode veće količine mlijeka od Simentalske i Smedje pasmine.

SAS primjer okomitog pravokutnog grafikona

```
TITLE 'Proizvodnja mlijeka po pasminama tijekom laktacija';
PROC SGPLOT DATA=dipl.Hpa_podaci;
VBOX my /CATEGORY=laktacija GROUP=pasmina DATALABEL;
KEYLEGEND / LOCATION=INSIDE POSITION=TOPRIGHT ACROSS=1;
YAXIS LABEL='Kolicina mlijeka (kg)';
RUN;
```

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. Korištena je VBOX naredba čime radimo okomiti pravokutni grafikon. Za prikaz podataka korištena je numerička varijabla (*my*). Opcija CATEGORY omogućuje nam prikaz podataka po kategorijama, kategorizirali smo pomoću varijable (*laktacija*), tako da dobijemo pravokutni prikaz za svaku laktaciju. GROUP opcijom grupiramo prikaz podataka, grupirali smo pomoću varijable (*pasmina*), tako da nam se prikažu pravokutnici prema pasminama. Opcijom DATALABEL označavamo ispis vrijednosti koje se nalaze van pravokutnog grafikona, odnosno označavamo ekstremna opažanja. Unutar naredbe KEYLEGEND određujemo način na koji će nam se prikazivati legenda te gdje će se nalaziti. LOCATION opcija koristi se za promjenu lokaciju legende, dakle nalazi li se legenda unutar ili van grafikona. Opcija POSITION govori nam o smještaju grafikona, smještaj može biti desno, lijevo i na sredini. Opcijom ACROSS korigiramo prikaz koloni u legendi, ACROSS 1 prikazuje nam varijable u jednoj koloni.



Slika 12. Prikaz Pravokutnog grafikona SGPLOT procedurom

Objašnjenje grafikona: Na slici 12 prikazan je okomiti pravokutni grafikon. Na ovom grafikonu prikazali smo prosječnu proizvodnju mlijeka po pasminama tijekom laktacija. Na Y osi prikazana je količina proizvedenog mlijeka izražena u kilogramima, dok su na X osi prikazane laktacije od 1 do 4. Na legendi je prikazana klasifikacija pasmina prema bojama. Budući da smo podatke grupirali prema pasminama, plavom bojom prikazana je Simentalska pasmina, crvenom bojom Holstein pasmina, dok su zelenom bojom prikazane Smedje pasmine. Zahvaljujući kategorizaciji podaci su nam prikazani prema laktacijama od 1 do 4 ovisno o laktaciji u kojoj su se krave nalazile. Iz ovog grafikona možemo vidjeti da je najviša proizvodnja mlijeka u 2. i 3. laktaciji, dok je najniža proizvodnja u 4. laktaciji, te da Holstein goveda proizvode veće količine mlijeka od Simentalske i Smedje pasmine.

3.4. Grafikoni distribucija kategoričkih varijabli

3.4.1. Stupčasti grafikon (eng. Bar chart)

Prema Delwiche i Slaughter (2012) stupčasti grafikoni se koriste za prikaz distribucije kategoričke varijable gdje je dužina svakog stupca proporcionalna broju opažanja u pojedinoj kategoriji. Matange i Heath (2011) tvrde da ukoliko posjedujemo podatke o kojima imamo puno informacija svaki stupac može se prikazati kao skup manjih grupa stupaca. Kao i kod pravokutnih grafikona, velika količina ugrađenih naredbi omogućuje nam različite prikaze okomitih i vodoravnih stupčastih grafikona.

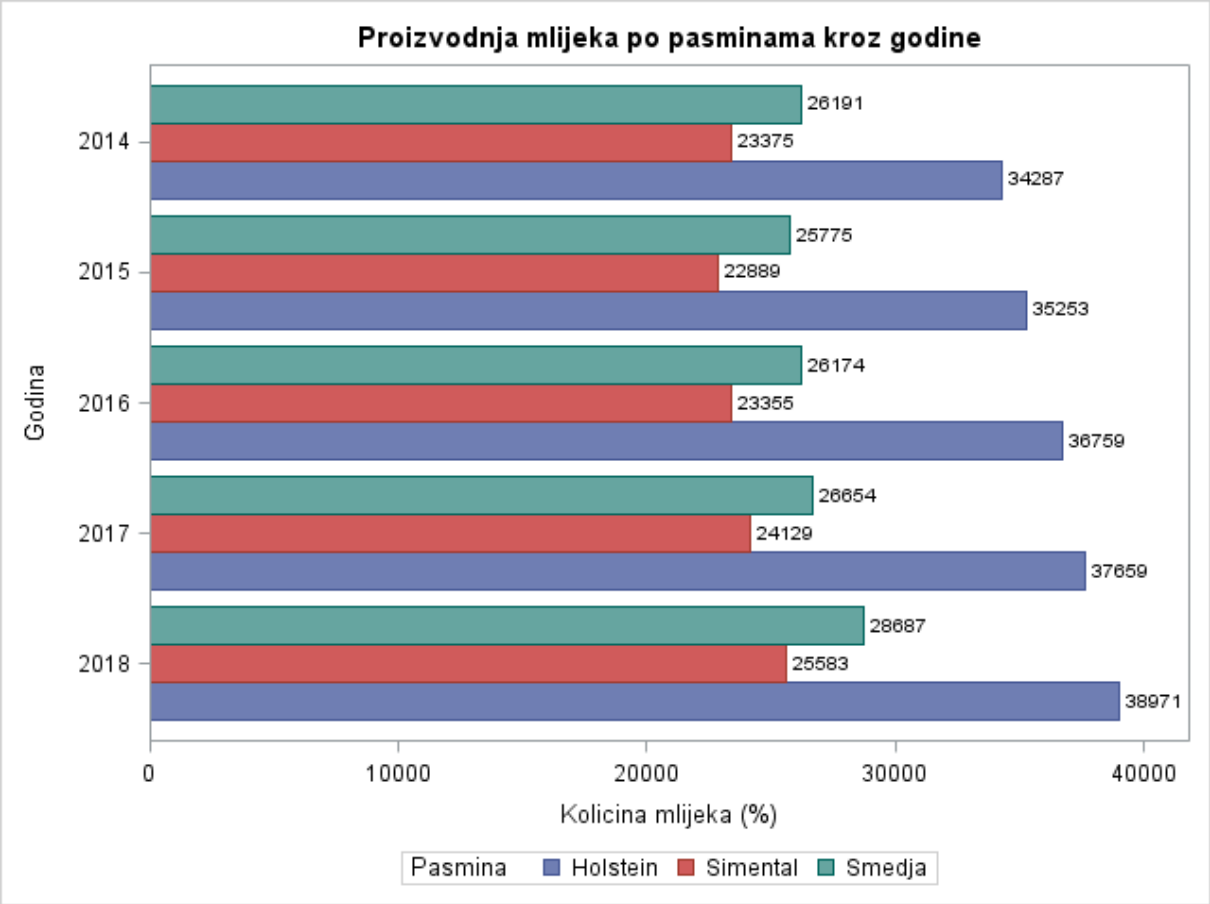
Sintaksa SAS programa prema Delwiche i Slaughter (2012)

```
PROC SGPLOT;
VBAR/HBAR naziv-varijable </ opcije>;
RUN;
```

SAS primjer za vodoravni stupčasti grafikon

```
TITLE 'Proizvodnja mlijeka po pasminama kroz godine';
PROC SGPLOT DATA=dipl.Hpa_podaci;
HBAR godina / GROUP=pasmina RESPONSE=my GROUPDISPLAY=cluster DATALABEL;
XAXIS LABEL='Kolicina mlijeka (kg)';
YAXIS LABEL='Godina';
RUN;
```

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. Korištena je HBAR naredba čime radimo vodoravni stupčasti grafikon. Za prikaz je korištena kategorička varijabla (*Godina*). Opcijom GROUP grupiramo prikaz podataka, grupirali smo pomoću varijable (*pasmina*), tako da nam se stupci prikažu prema pasminama. Opcijom RESPONSE prikazujemo opažanja varijable prema stupcima, mi smo prikazali vrijednosti količine mlijeka (*my*). Opcijom GROUPDISPLAY određujemo na koji će način biti prikazane grupe, u našem primjeru prikazan je opcijom CLUSTER čime smo postigli da su skupljeni stupci prema godini. Opcijom DATALABEL označavamo ispis vrijednosti, mi smo prikazali vrijednosti količine mlijeka (*my*) kraj vrha svakog stupca. Unutar naredbe YAXIS opcijom LABEL definiramo naziv varijable koje će se nalaziti na Y osi.



Slika 13. Prikaz vodoravnog Stupčastog grafikona SGPLOT procedurom

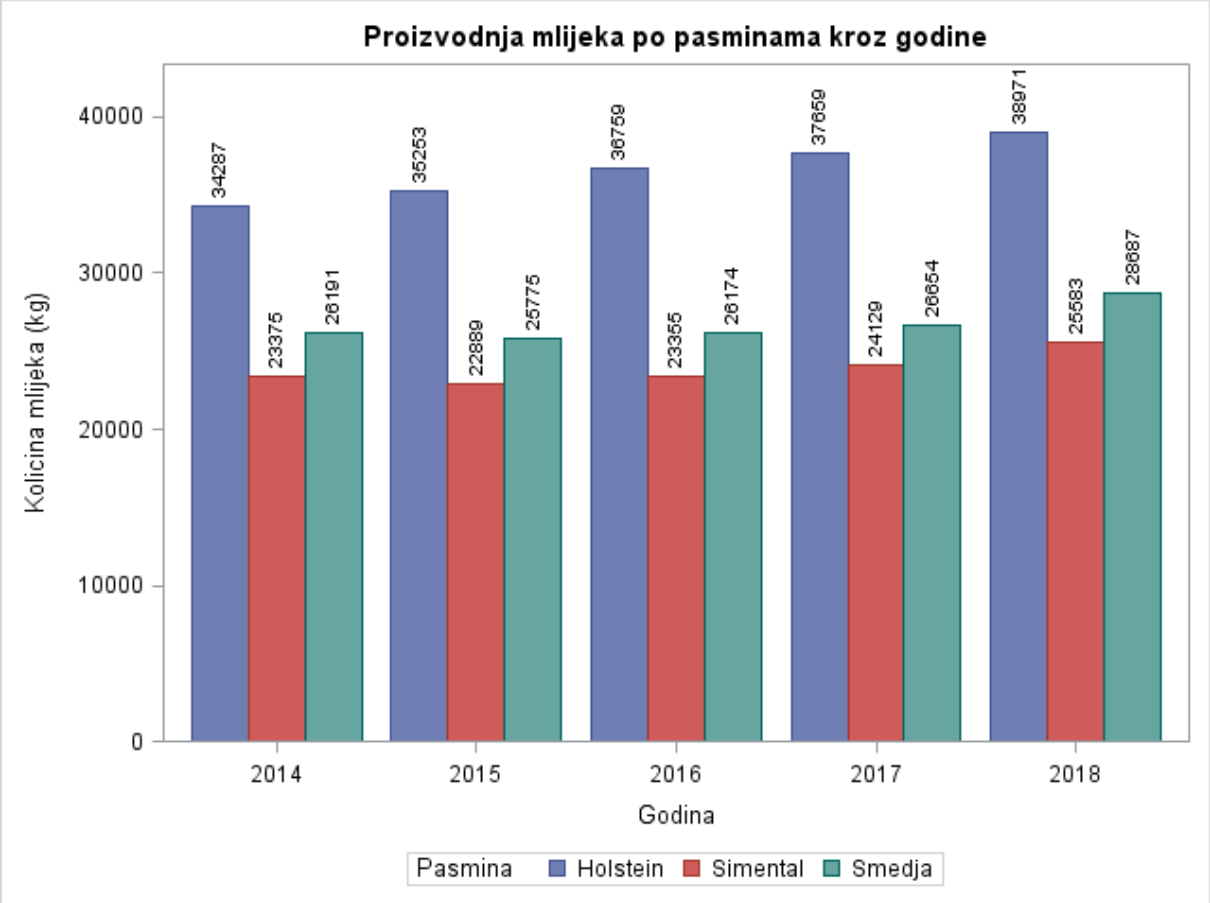
Objašnjenje grafikona: Na slici 13 prikazan je vodoravni stupčasti grafikon. Na ovom grafikonu prikazali smo proizvodnju mlijeka po pasminama između 2014. i 2018. godine. Budući da smo podatke grupirali prema pasminama, plavom bojom je prikazana Holstein pasmina, crvenom bojom Simentalska pasmina, dok je zelenom bojom prikazana Smedja pasmina. Na Y osi prikazane su godine, dok je na X osi prikazana količina proizvedenog mlijeka izražena u kilogramima. Na legendi je prikazana klasifikacija pasmina prema bojama. Podaci ispisani kraj stupaca govore nam o količini proizvedenog mlijeka po pasmini. Isto tako možemo vidjeti da se količina proizvedenog mlijeka postepeno povećava kod Holstein pasmine, dok se kod Simental goveda i Smedjih pasmina povećanje proizvodnje javlja od 2016. godine.

SAS primjer za okomiti stupčasti grafikon

```
TITLE 'Proizvodnja mlijeka po pasminama kroz godine';
PROC SGPLOT DATA=dipl.Hpa_podaci;
VBAR godina / GROUP=pasmina RESPONSE=my GROUPDISPLAY=cluster DATALABEL;
```

YAXIS LABEL='Kolicina mlijeka (kg)';
RUN;

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. Korištena je VBAR naredba čime radimo okomiti stupčasti grafikon. Za prikaz je korištena kategorička varijabla (*Godina*). Opcijom GROUP grupiramo prikaz podataka, grupirali smo pomoću varijable (*pasmina*), tako da nam se stupci prikažu prema pasminama. Opcijom RESPONSE prikazujemo opažanja varijable prema stupcima, mi smo prikazali vrijednosti količine mlijeka (*my*). Opcijom GROUPDISPLAY određujemo na koji će način biti prikazane grupe, u našem primjeru prikazan je opcijom CLUSTER čime smo postigli da su skupljeni stupci prema godini. Opcijom DATALABEL označavamo ispis vrijednosti, mi smo prikazali vrijednosti količine mlijeka (*my*) kraj vrha svakog stupca. Unutar naredbe XAXIS opcijom LABEL definiramo naziv varijable koje će se nalaziti na Y osi.



Slika 14. Prikaz okomitog stupčastog grafikona SGPLOT procedurom

Objašnjenje grafikona: Na slici 14 prikazan je okomiti stupčasti grafikon. Na ovom grafikonu prikazali smo proizvodnju mlijeka po pasminama između 2014. i 2018. godine. Budući da smo podatke grupirali prema pasminama, plavom bojom je prikazana Holstein

pasmina, crvenom bojom Simentalska pasmina, dok je zelenom bojom prikazana Smedja pasmina. Na Y osi prikazana je količina proizvedenog mlijeka izražena u kilogramima, dok su na X osi prikazane godine. Na legendi je prikazana klasifikacija pasmina prema bojama. Podaci ispisani kraj stupaca govore nam o količini proizvedenog mlijeka po pasmini. Isto tako možemo vidjeti da se količina proizvedenog mlijeka postepeno povećava kod Holstein pasmine, dok se kod Simental goveda i Smedjih pasmina povećanje proizvodnje javlja od 2016. godine.

3.4.2. Točkasti grafikon (eng. Dot plot)

Prema Matange i Heath (2011) točkasti grafikon prikazuje skup opažanja varijable po kategoriji ili zbroju frekvencije kategorije te su vrlo učinkoviti za usporedbu velikog broja vrijednosti unutar kategorija. Nadalje Matange i Heath (2011) tvrde da svako opažanje označava jednu kategoriju tj. vrijednost grupe.

Sintaksa SAS programa prema Delwiche i Slaughter (2012)

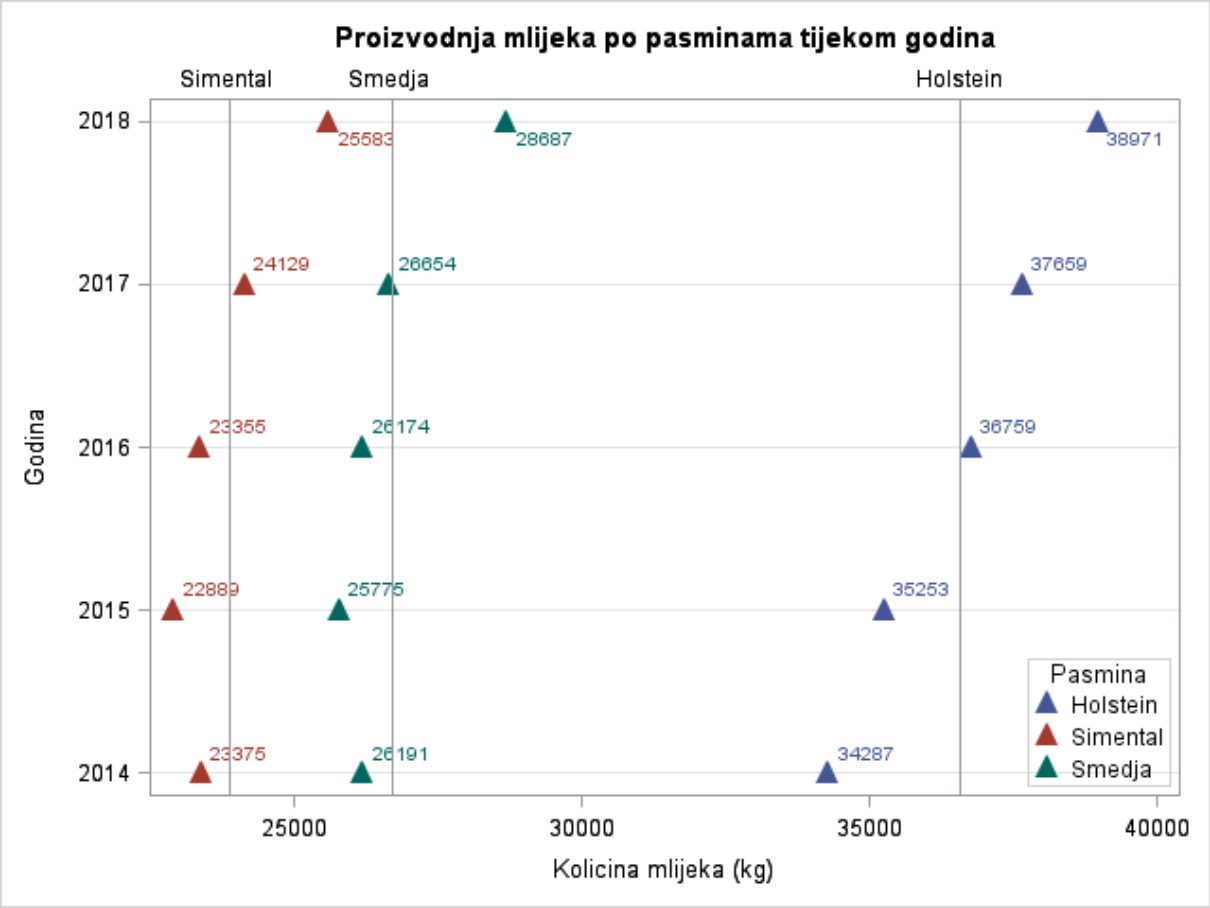
```
PROC SGPLOT;
DOT naziv varijable </ opcije>;
RUN;
```

SAS primjer točkastog grafikona

```
TITLE 'Proizvodnja mlijeka po pasminama tijekom godina';
PROC SGPLOT DATA=dipl.Hpa_podaci;
FORMAT godina best6.;
DOT godina / RESPONSE=my GROUP=pasmina
MARKERATTRS=(SYMBOL=TRIANGLEFILLED SIZE=12px) DATALABEL;
REFLINE 23866 36586 26696/ LABEL=("Simmental" "Holstein" "Brown") AXIS=x;
KEYLEGEND / LOCATION=INSIDE ACROSS=1;
XAXIS LABEL='Kolicina mlijeka (kg)';
YAXIS DISCRETEORDER=FORMATTED REVERSE;
RUN;
```

Objašnjenje: Naredba TITLE daje nam naziv grafikonu. Korištena je DOT naredba. Za prikaz podataka koristimo kategoričku varijablu (*godina*). Naredbom FORMAT određujemo kojim formatom želimo prikazati varijablu, mi smo formatirali prikaz varijable (godina). Opcijom RESPONSE prikazujemo opažanja varijable na grafikonu, mi smo prikazali

vrijednosti količine mlijeka (*my*) Opcijom GROUP grupiramo prikaz podataka, grupirali smo pomoću varijable (*pasmina*), tako da nam se opažanja prikažu prema pasminama. MARKERATTRS opcijom mijenjamo izgled simbola opcijom SYMBOL, odabrali smo prikaz ispunjenog trokuta, dok smo opcijom SIZE promijenili veličinu opažanja, tako da nam budu jasno prikazana. Opcijom DATALABEL označavamo ispis vrijednosti, mi smo prikazali vrijednosti količine mlijeka (*my*) kraj svakog opažanja. Naredbom REFLINE određujemo vrijednosti referentnih linija, LABEL opcijom određujemo naziv referentne linije, dok pomoću AXIS opcije određujemo na kojoj osi se prikazuje referentna linija. Unutar naredbe KEYLEGEND pomoću opcija LOCATION i ACROSS određujemo način prikazivanja legende te gdje i na koji način prikazati. LOCATION opcija koristi se za promjenu lokaciju legende, dakle nalazi li se legenda unutar ili van grafikona. Opcijom ACROSS korigiramo prikaz koloni u legendi, ACROSS 1 prikazuje nam varijable u jednoj koloni. Unutar YAXIS naredbe DISCRETEORDER opcijom određujemo način prikaza varijable, u našem prikazu želimo da nam se podaci prikazuju kako su formatirani. REVERSE opcijom okrećemo prikaz varijable, u našem primjeru želimo da nam se podaci na Y osi pokazuju od najmanjeg prema najvećem.



Slika 15. Prikaz točkastog grafikona SGPLOT procedurom

Objašnjenje grafikona: Na slici 15 prikazan je točkasti grafikon. Na ovom grafikonu prikazali smo proizvodnju mlijeka po pasminama u između 2014. i 2018. godine. Budući da smo podatke grupirali prema pasminama, plavom bojom prikazana je Holstein pasmina, crvenom bojom Simentalska, dok je zelenom bojom prikazana Smedja pasmina. Podaci ispisani kraj opažanja govore nam o količini proizvedenog mlijeka. Osim samih opažanja, na grafikonu su prikazane referentne linije svih pasmina. Na Y osi prikazane su godine, dok je na X osi prikazana količina proizvedenog mlijeka izražena u kilogramima. Na legendi je prikazana klasifikacija pasmina prema bojama, te je prikazana okomito zbog korištenja ACROSS opcije. Iz ovog grafikona možemo vidjeti da je prosječna količina proizvedenog mlijeka porasla kod svih pasmina. Kod Holstein goveda je vidljiv najveći porast proizvodnje, dok je kod Simental i Smedjih pasmina vidljiv blagi porast u proizvodnosti nakon 2016. godine.

4. GPLOT procedura

Cartier (2002) tvrdi da je GPLOT procedura jedna od pet glavnih procedura SAS/GRAPH paketa SAS software-a. Nadalje prema Cartier (2002), SAS/GRAPH paket je koristan alat za izradu vrlo kvalitetnih grafikona. Prema Li (2015) vrste grafikona koje možemo izraditi koristeći ovu proceduru su raspršeni, višeslojnih, logaritamski i mjehuričasti grafikon. Osim navedenih vrsta grafikona mogu se napraviti i druge vrste grafikona korištenjem SYMBOL naredbe. SYMBOL naredbu je najbolje koristiti za prikaz dužih nizova podataka, interpolaciju između dvije točke na grafikonu, ekstrapolaciju između postojećih podataka s prikazom regresijskih linija i razine značajnosti (SAS Institute Inc., 2016). Cartier (2002) tvrdi da određivanjem interpolacije između točaka na grafikonu pomoću INTERPOL= opcije može se nacrtati linearna regresija ili nacrtati linijski grafikon koji prolazi kroz točke na grafikonu te nam na taj način prikazuje trendove naših podataka. Prema Li (2015) GPLOT procedurom možemo preklapati najviše dvije plot naredbe.

Sintaksa SAS programa za GPLOT proceduru prema Li (2015)

PROC GPLOT <opcije>;
PLOT/BUBBLE zahtjevi-za-plot </ opcije>;
PLOT2/BUBBLE2 zahtjevi-za-plot </ opcije>;
RUN;

4.1. Raspršeni grafikon (eng. Scatter plot)

Cartier (2002) tvrdi da su za stvaranje dvodimenzionalnog raspršenog grafikona na koordinatnom sustavu GPLOT procedurom koristeći PLOT naredbu potrebne dvije varijable te je moguće dodati opcionalnu treću varijablu kojom se definiraju podgrupe podataka.

Sintaksa SAS programa za GPLOT proceduru prema SAS Institute Inc. (2016)

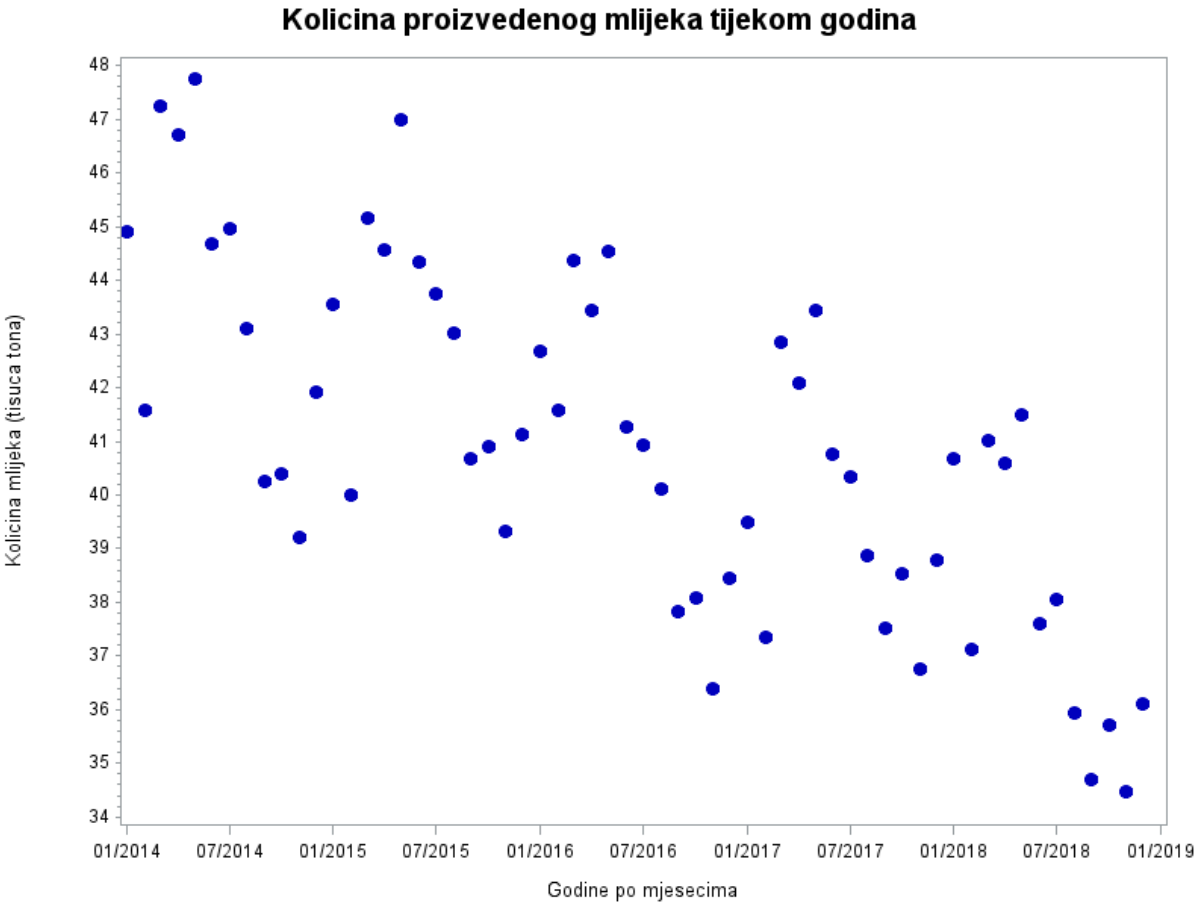
PROC GPLOT DATA=SAS-data-set ;
PLOT zavisna-varijabla* nezavisna-varijabla </ opcije>;
RUN;

SAS primjer raspršenog grafikona koristeći GPLOT proceduru

AXIS1 LABEL=('Godine' H=1);

```
AXIS2 LABEL=(A=90 J=C 'Kolicina mlijeka (tisuca tona)' H=1);
TITLE 'Kolicina proizvedenog mlijeka tijekom godina';
SYMBOL1 VALUE=dot COLOR='dark blue' H=1.5 W=1.5;
PROC GPLOT DATA=dipl.Eurostat;
PLOT value*TIME / HAXIS=AXIS1 VAXIS=AXIS2;
RUN;
```

Objašnjenje: Prije same GPLOT procedure koristimo AXIS1 i AXIS2 naredbe za prilagođavanje prikaza naziva varijabli koje se nalaze na X i Y osima. LABEL opcijom dajemo naziv varijabli koja se nalazi na X i Y osi, dok H opcijom određujemo veličinu slova. Opcijom A korigiramo kut prikaza naziva varijable u stupnjevima, mi želimo da nam se naziv prikaže okrenut za 90 stupnjeva. Opcijom J određujemo gdje će se nalaziti naziv varijable, u našem prikazu želimo da nam se natpis nalazi u sredini (center). Naredba TITLE daje nam naziv grafikonu. Unutar naredbe SYMBOL1 opcijom VALUE određujemo kojim će simbolima biti označene vrijednosti, u našem prikazu korišten je prikaz točkom. COLOR opcijom određujemo boju kojom želimo prikazati opažanja, mi smo prikazali tamno plavom bojom. Opcijama H i W određujemo veličinu prikaza točaka, H se koristi za visinu opažanja, dok se W koristi za širinu opažanja. Za prikaz podataka koristimo nezavisnu (*TIME*) i zavisnu varijablu (*value*). Unutar PLOT naredbe HAXIS i VAXIS opcijama dodajemo korigirane vrijednosti na X i Y osima. HAXIS opcijom dodajemo modificirani prikaz naziva varijable na X osi. VAXIS opcijom dodajemo modificirani prikaz naziva varijable na Y osi.



Slika 16. Prikaz raspršenog grafikona GPLOT procedurom

Objašnjenje grafikona: Na slici 16 prikazan je raspršeni grafikon. Na ovom grafikonu prikazali smo količinu proizvedenog mlijeka u Republici Hrvatskoj po mjesecima između 2014. i 2019. godine. Na Y osi prikazana je količina proizvedenog mlijeka izražena u tisućama tona, dok su na X osi prikazane godine po mjesecima. U legendi je prikazana vrijednost varijable Količina proizvedenog mlijeka u RH koja je prikazana na grafikonu. Plavim točkama prikazani su podatci za količinu proizvedenog mlijeka u RH između 2014. i 2019. godine. Možemo vidjeti da je količina proizvedenog mlijeka svake godine najviša početkom godine te se smanjuje prema sredini godine i ponovno raste prema kraju godine.

4.2. Mjehuričasti grafikona (eng. Bubble plot)

U SAS Institute Inc. (2016) navodi se da se mjehuričastim grafikonom prikazuju vrijednosti tri varijable u koordinatnom sustavu crtanjem krugova različitih veličina. U istome radu tvrdi se da dvije varijable određuju mjesto točke u sustavu dok treća varijabla određuje veličinu kruga. Nadalje u istome radu, prema zadanim postavkama mjehuričastog grafikona prilagođava osi koordinatnog sustava tako da se mogu prikazati i najveće i najmanje vrijednosti.

Sintaksa SAS programa za mjehuričasti grafikon prema SAS Institute Inc. (2016)

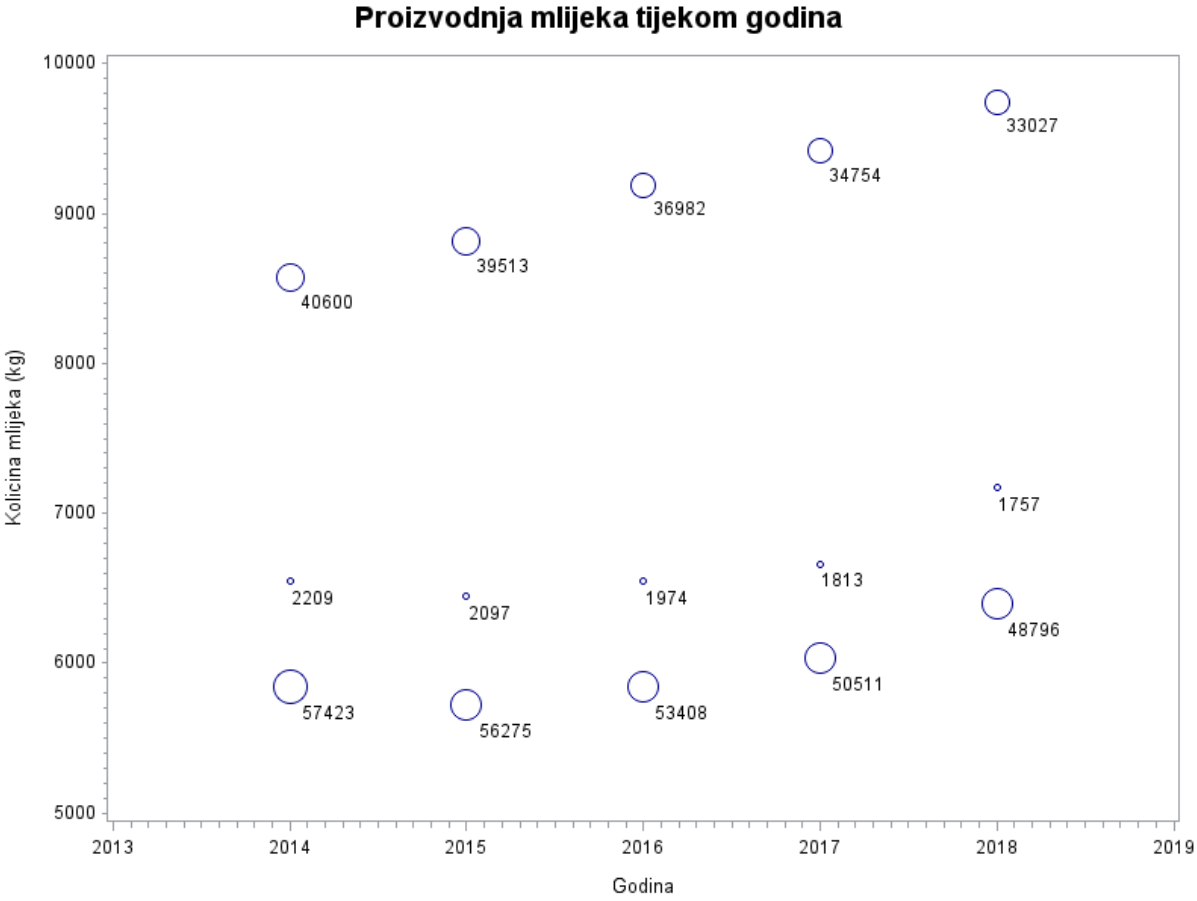
```
PROC GPLOT ;
BUBBLE zavisna-varijabla*nezavisna-varijabla = veličina-mjehurića </ opcije>;
RUN;
```

SAS primjer za mjehuričasti grafikon

```
GOPTIONS BORDER;
AXIS1 ORDER=(2013 to 2019 by 1)
            LABEL=('Godina' H=1);
AXIS2 LABEL=(A=90 J=C 'Kolicina mlijeka (kg)' H=1);
TITLE 'Proizvodnja mlijeka tijekom godina';
PROC GPLOT DATA=dipl.Hpa_podaci2;
BUBBLE my*godina = stado / BSCALE=pasmina BCOLOR=darkblue BLABEL
HAXIS=AXIS1 VAXIS=AXIS2;
RUN;
```

Objašnjenje: Unutar naredbe GOPTIONS opcijom BORDER prikazujemo okvir oko grafikona. Prije same GPLOT procedure koristimo AXIS1 i AXIS2 naredbe za prilagođavanje prikaza naziva varijabli koje se nalaze na X i Y osima. Pomoću ORDER i BY opcije određujemo redoslijed prikaza vrijednosti varijable, u našem slučaju želimo da se podaci prikazuju od 2013 do 2019, te da se povećavaju za 1. LABEL opcijom dajemo naziv varijabli koja se nalazi na X i Y osi, dok H opcijom određujemo veličinu slova. Opcijom A korigiramo kut prikaza naziva varijable u stupnjevima, mi želimo da nam se naziv prikaže okrenut za 90 stupnjeva. Opcijom J određujemo gdje će se nalaziti naziv varijable, u našem prikazu želimo da nam se natpis nalazi u sredini (center). Naredba TITLE daje nam naziv grafikonu. Unutar naredbe SYMBOL1 opcijom VALUE određujemo kojim će simbolima biti označene vrijednosti, u našem prikazu korišten je prikaz točkom. COLOR opcijom određujemo boju kojom želimo prikazati opažanja, mi smo prikazali tamno plavom bojom. Opcijama H i W određujemo veličinu prikaza točaka, H se koristi za visinu opažanja, dok se W koristi za širinu opažanja. U BUBBLE naredbi za prikaz podataka koristimo nezavisnu (*godina*) i zavisnu varijablu (*my*) i varijablu koja nam govori o veličini mjehurića (*stado*). Opcijom BSCALE određujemo način prikaza podataka, u našem primjeru želimo da se podaci prikazuju prema varijabli (*pasmina*). Opcijom BCOLOR određujemo boju mjehurića, prikazali smo mjehuriće tamnoplavom bojom. Opcijom BLABEL dodajemo vrijednosti varijable uz opažanja, mi smo prikazali vrijednost varijable (*stado*) kraj mjehurića. HAXIS i VAXIS opcijama dodajemo

korigirane vrijednosti na X i Y osima. HAXIS opcijom dodajemo modificirani prikaz naziva varijable na X osi. VAXIS opcijom dodajemo modificirani prikaz naziva varijable na Y osi.



Slika 17. Prikaz mjehuričastog grafikona GPLOT procedurom

Objašnjenje grafikona: Na slici 17 prikazan je mjehuričasti grafikon. Na ovom grafikonu prikazali smo proizvodnju mlijeka po pasminama između 2014. i 2018. godine. Na grafikonu su prikazani podaci o Holstein, Smedjoj i Simentalskoj pasmini. Mjehurići su prikazani prema pasminama, ali nažalost nisu obojani različitim bojama, tako da iz grafikona ne možemo saznati o kojim pasminama se radi osim ako nismo dobro upućeni u podatke. Podaci ispisani kraj mjehura govore nam o broju grla u kontroli mliječnosti, što nam različite veličine mjehura slikovno prikazuju svojom veličinom. Iz ovog grafikona možemo vidjeti da je prosječna količina proizvedenog mlijeka porasla kod svih pasmina.

4.3. Višeslojni grafikon (eng. Overlay plot)

U SAS Institute Inc. (2016) navedeno je da unutar PLOT naredbe postoji OVERLAY opcija koja nam omogućuje prikaz obje linije kroz podatke na istom grafikonu. Nadalje u istome radu, ukoliko ne definiramo OVERLAY opciju dobit ćemo dva zasebna grafikona s unesenim vrijednostima.

Sintaksa SAS programa za višeslojni grafikon prema Cartier (2002)

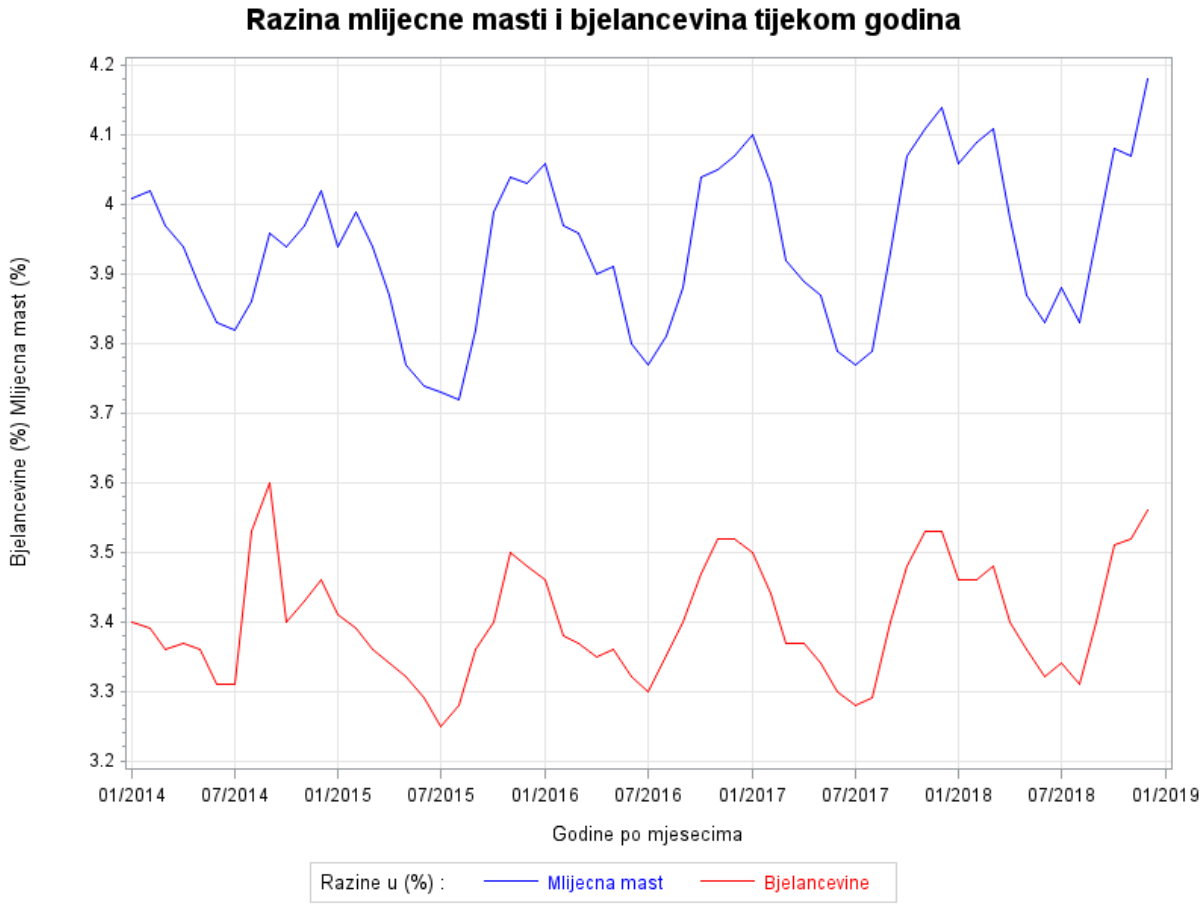
```
PROC GPLOT;
PLOT Y1*X Y2*X Y3*X / OVERLAY;
```

SAS primjer za višeslojni grafikon koristeći GPLOT proceduru

```
AXIS1 LABEL=('Godine po mjesecima' H=1);
AXIS2 LABEL=(A=90 J=c "Bjelancevine (%) Mlijecna mast (%)" H=1);
LEGEND1 POSITION=(OUTSIDE BOTTOM) FRAME
                LABEL= ('Razine u (%) :' H=1)
                VALUE= (T=1 C=blue 'Mlijecna mast' H=1
                                T=2 C=red 'Bjelancevine' H=1);
GOPTIONS BORDER COLORS=(blue red) INTERPOL=join;
TITLE 'Razina mlijecne masti i bjelancevina tijekom godina';
PROC GPLOT DATA=dipl.Eurostat;
PLOT mm*time prot*time / HAXIS=axis1 VAXIS=axis2 LEGEND=legend1 OVERLAY
GRID;
RUN;
```

Objašnjenje: Prije same GPLOT procedure koristimo AXIS1 i AXIS2 naredbe za prilagođavanje prikaza naziva varijabli koje se nalaze na X i Y osima. Prije same GPLOT procedure koristimo AXIS1 i AXIS2 naredbe za prilagođavanje prikaza naziva varijabli koje se nalaze na X i Y osima. Pomoću ORDER i BY opcije određujemo redoslijed prikaza vrijednosti varijable, u našem slučaju želimo da se podaci prikazuju od 2013 do 2019, te da se povećavaju za 1. LABEL opcijom dajemo naziv varijabli koja se nalazi na X i Y osi, dok H opcijom određujemo veličinu slova. Opcijom A korigiramo kut prikaza naziva varijable u stupnjevima, mi želimo da nam se naziv prikaže okrenut za 90 stupnjeva. Opcijom J određujemo gdje će se nalaziti naziv varijable, u našem prikazu želimo da nam se natpis nalazi u sredini. Naredbom LEGEND1 radimo legendu Opcijom POSITION definiramo njenu

poziciju, što BOTTOM i OUTSIDE opcijama i činimo tako da je legenda smještena ispod grafikona, i van njega. Opcijom FRAME dajemo legendi okvir. Opcijom Value definiramo vrijednosti koje će se nalaziti unutar legende. T opcijom definiramo koja će po redu biti vrijednost, u našem primjeru koja je vrijednost napisana prva, a koja druga. Opcijom C definiramo boju kojom će biti napisana vrijednost u legendi. Unutar naredbe GOPTIONS opcijom BORDER prikazujemo okvir oko grafikona. Opcijom COLORS definiramo kojim će nam bojama biti prikazane vrijednosti na grafikonu. Opcijom INTERPOL definiramo način kako su opažanja prikazana, u našem primjeru želimo da su opažanja povezana te zato koristimo JOIN opciju. Naredba TITLE daje nam naziv grafikonu. Unutar naredbe SYMBOL1 opcijom VALUE određujemo kojim će simbolima biti označene vrijednosti, u našem prikazu korišten je prikaz točkom. COLOR opcijom određujemo boju kojom želimo prikazati opažanja, mi smo prikazali tamno plavom bojom. Opcijama H i W određujemo veličinu prikaza točaka, H se koristi za visinu opažanja, dok se W koristi za širinu opažanja. U BUBBLE naredbi za prikaz podataka koristimo nezavisnu (*godina*) i zavisne varijable (*mm, prot*). HAXIS i VAXIS opcijama dodajemo korigirane vrijednosti na X i Y osima. HAXIS opcijom dodajemo modificirani prikaz naziva varijable na X osi. VAXIS opcijom dodajemo modificirani prikaz naziva varijable na Y osi. LEGEND opcijom dodajemo izmijenjeni prikaz legende. OVERLAY opcijom preklapamo dva linijska grafikona na istom grafikonu. Grid opcijom dodajemo linije koordinatnog sustava.



Slika 18. Prikaz višeslojnog linijskog grafikona GPLOT procedurom

Objašnjenje grafikona: Na slici 18 prikazan je višeslojni linijski grafikon. Na ovom grafikonu prikazali smo razinu mliječne masti i bjelančevina između 2014. i 2019. godine. Budući da smo na istom grafikonu prikazali dvije zavisne varijable, one su prikazane različitim bojama. Na Y osi prikazana je razina bjelančevina i mliječne masti u mlijeku izražena u postotcima, dok su na X osi prikazane godine po mjesecima. Na legendi su prikazane vrijednosti razina prikazane na grafikonu izražene u postotcima. Razina mliječne masti u mlijeku prikazana je plavom bojom, dok je razina bjelančevina u mlijeku prikazana crvenom bojom. Iz ovog grafikona možemo zaključiti da su se najviše vrijednosti mliječne masti neznatno povećale tijekom godina, dok je razina bjelančevina ostala gotovo ista. Isto tako možemo vidjeti da je razina i bjelančevina i mliječne masti najviša početkom godine te da se smanjuje do 6. mjeseca kada je najniža, nakon čega slijedi rast prema kraju godine.

5. Usporedba SGPLOT i GPLOT procedura

SAS® nam omogućuje prikaz raznih statističkih grafičkih procedura. SG procedure puno su snažnije tako da uz minimalno kodiranje možemo prikazati kompleksnije podatke te je isto tako omogućena lakša automatizacija i prilagodba podataka grafičkom prikazu (Amruthnath 2013; Canchola i sur. 2018).

Prema Matange i Heath (2011) SG procedure funkcioniraju u sklopu ODS Graphics sustava te koriste GTL (graph template language) za stvaranje grafičkog prikaza. Nadalje Matange i Heath (2011) tvrde da se zbog korištenja GTL-a SG procedure smatra "template-based graphics", dok se SAS/GRAPH procedura smatra "device-based graphics".

Glavne razlike SGPLOT i GPLOT procedura prema Matange i Heath (2011)

- SG procedure stvaraju output u formatima koji se koriste u industriji poput: PNG, JPG, PDF, PS itd., dok GPLOT procedure stvaraju GRSEG output. Pomoću SGPLOT procedure se može na puno jednostavniji način prikazati grafikon na ekranu, ali i za daljnju uporabu u znanosti i izdavaštvu za razliku od GPLOT procedure gdje se trebalo dodatno raditi kako bi se postigao sličan rezultat.
- SG procedure ne podržavaju tradicionalne GRAPH naredbe GOPTIONS, PATTERN, SYMBOL, AXIS, LEGEND. Međutim, pomoću drugih procedura, ODS Graphics ili ODS destination options moguće je postići jednake ili slične rezultate
- GOPTIONS naredba nije podržana kod SGPLOT procedure, osim za korištenje RESET opcije za resetiranje naslova i fusnoti. Dok su opcije za kontrolu outputa specificirane u ODS GRAPHICS naredbi.
- TITLE i FOOTNOTE naredbe nisu podržane kod SGPLOT procedure, osim za nekoliko opcija.
- PATTERN i SYMBOL naredbe nisu podržane kod SG procedura. Međutim, tipovi grafikona i atributa se specificiraju direktno u plot naredbama, tako da je uporaba tih atributa jednostavnija.
- AXIS naredba nije podržana. Međutim, PROC SGPLOT podržava četiri naredbe pomoću kojih možemo upravljati i koristiti osi sustava (XAXIS, YAXIS, X2AXIS, Y2AXIS).
- LEGEND naredba nije podržana, već PROC SGPLOT pomoću KEYLEGEND naredbe upravlja sadržajem, pozicijom i izgledom legende.

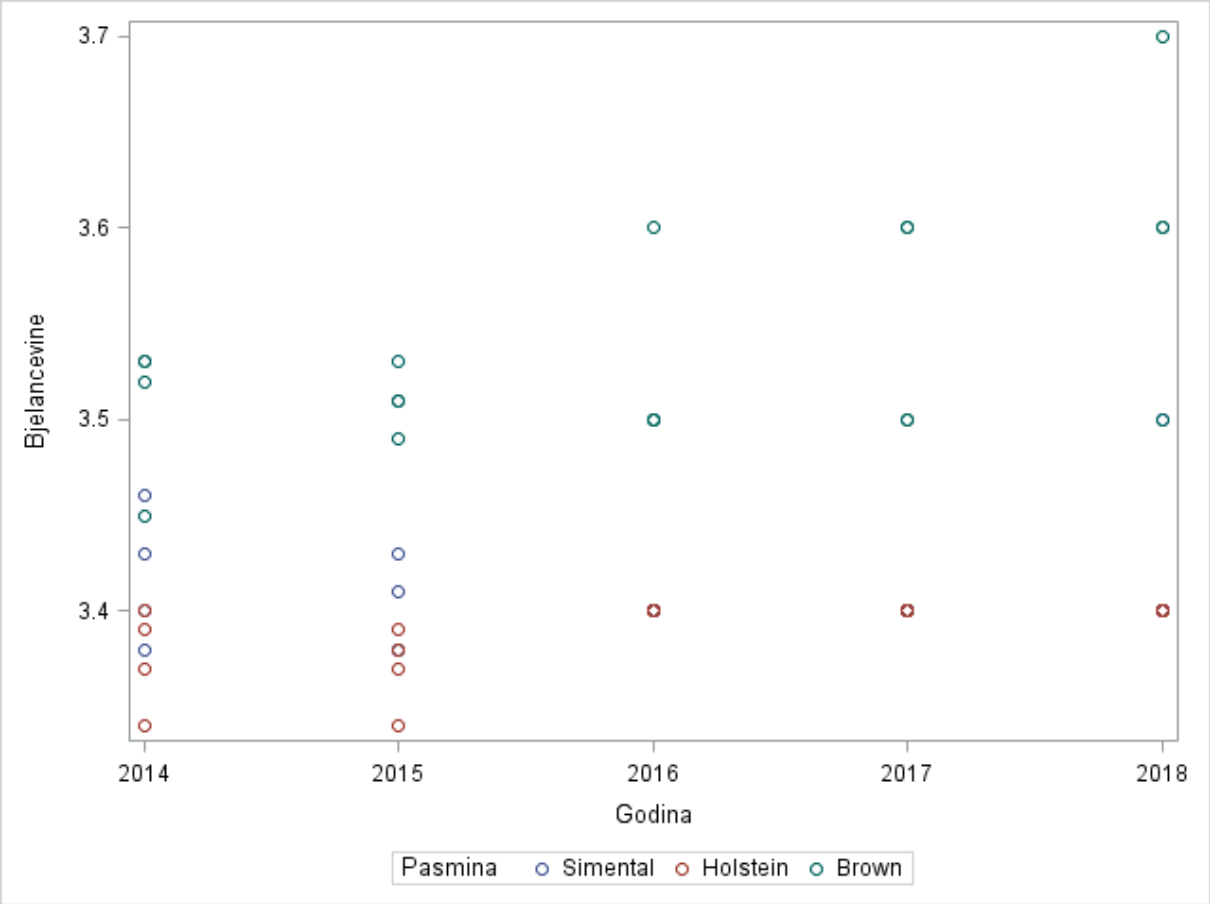
Prema Slaughter i Delwiche (2015) GPLOT procedura omogućuje nam stvoriti dvodimenzionalni grafikon koji tvori dvije ili više varijabli u koordinatnom sustavu. U istome radu tvrde da pomoću njega možemo prikazati različite vrste grafikona, ali je nažalost limitirana u mogućnostima proizvodnje raznih vrsta grafikona. Navode da može proizvesti samo raspršene, višeslojne, logaritamske i mjehuričaste plotove.

Isti autori navode kako SGPLOT procedurom možemo stvoriti neke osnovne grafikone poput raspršenih, linijskih i mjehuričastih grafikona, ali i grafikone prikaza modela i intervala pouzdanosti, distribucijske i grafikone kategoričkih varijabli. Isto tako u radu tvrde da možemo napraviti i sofisticiranije grafikone poput regresijskih, Loess-ove i B-Splines grafikona te ih sve možemo jednostavnije i ljepše prikazati.

Prema Slaughter i Delwiche (2015) izgled grafikona vrlo je lako korigirati i mijenjati korištenjem SGPLOT procedure i ODS GRAPHICS opcije koja nije potrebna ukoliko ne želimo promijeniti početne/zadane postavke. Isto tako u radu navode kako se kod GPLOT procedure izgled outputa svaki put korigira GOPTIONS opcijom, koja ostaje u funkciji nakon što ju prvi put unesemo.

Ranije u radu spomenuli smo razliku u kvaliteti izgleda grafikona koristeći SGPLOT i GPLOT procedure, pa ćemo pokušati prikazati razliku i na primjeru raspršenog grafikona. Za prikaz razlike u kvaliteti izgleda grafikona korišteni su podaci Hrvatske poljoprivredne agencije.

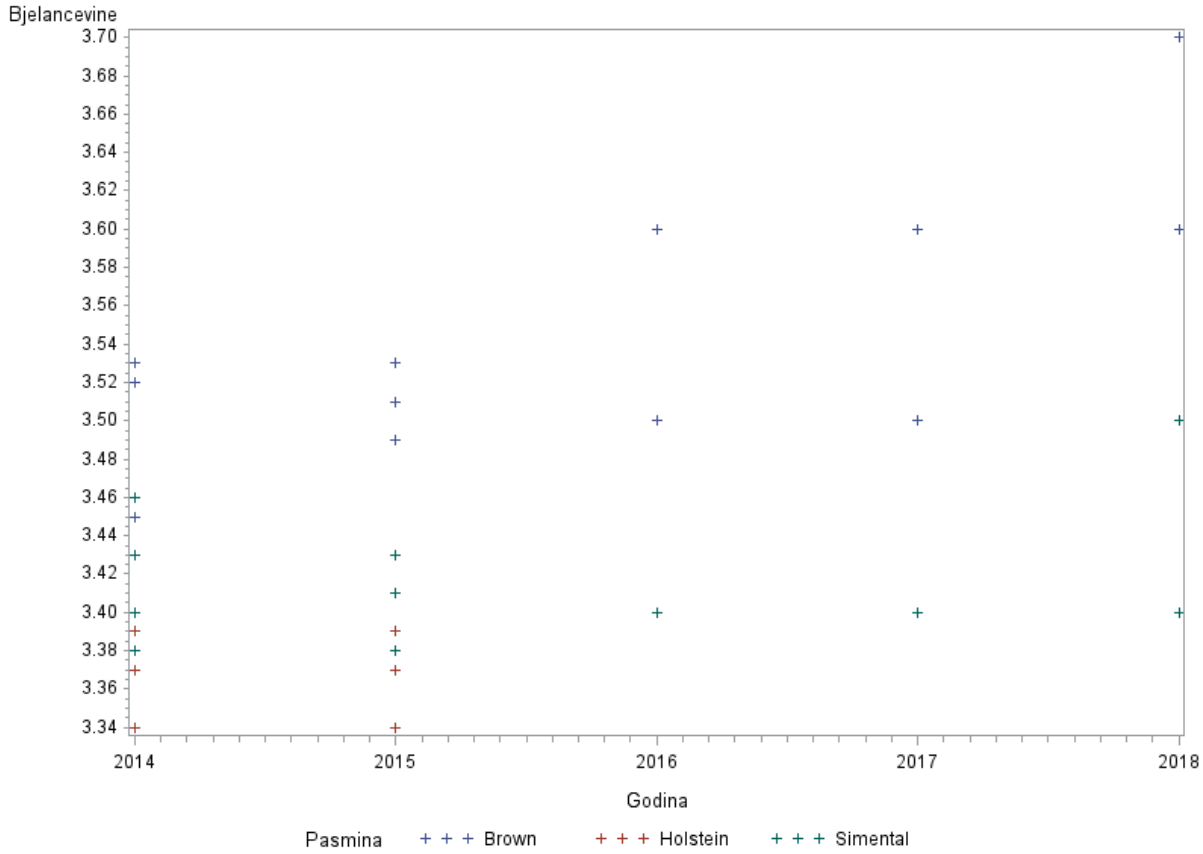
```
PROC SGPLOT DATA=dipl.Hpa_podaci;
SCATTER X=godina Y=bjelancevine / GROUP=pasmina;
WHERE PASMINA IN ("Brown", "Holstein", "Simental");
RUN;
```



Slika 19. Prikaz raspršenog grafikona SGPLOT procedurom

Za usporedbu smo napravili isti prikaz raspršenog grafikona koristeći GPLOT proceduru.

```
PROC GPLOT DATA=dipl.Hpa_podaci;
PLOT bjelancevine*godina=pasmina;
WHERE PASMINA IN ("Brown", "Holstein", "Simental");
RUN;
```

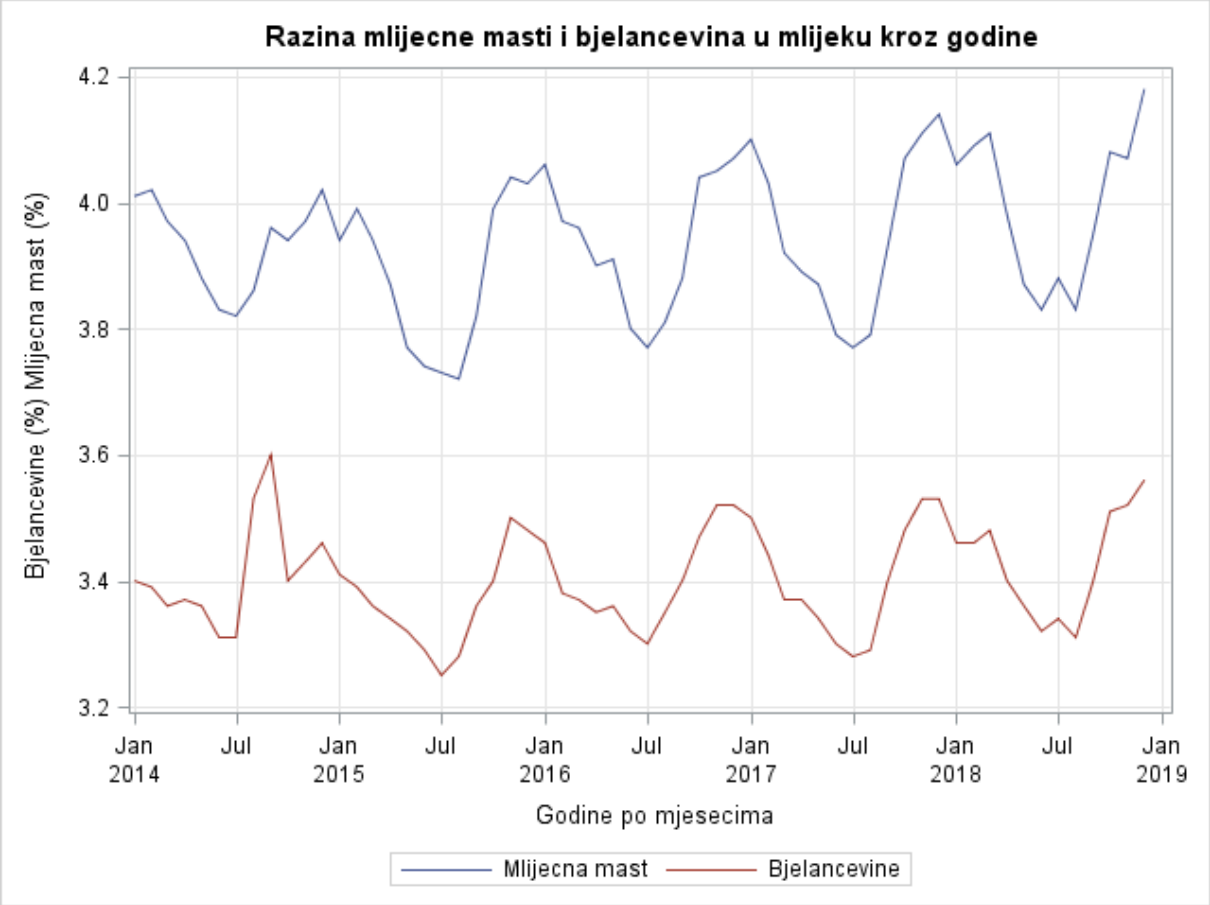


Slika 20. Prikaz raspršenog grafikona GPLOT procedurom

Glavna razlika koju odmah možemo primijetiti je način prikaza opažanja na grafikonu. Dok su u SGPLOT proceduri oblici i boje opažanja bolje vidljivi kod GPLOT procedure teže možemo raspoznati razliku između opažanja koja nam prikazuju razliku između pasmina. Ukoliko kod GPLOT procedure želimo promijeniti izgled opažanja u smislu boje, izgleda ili veličine to moramo učiniti pomoću SYMBOL naredbe u općim postavkama. Iduća stvar je razlika u prikazu naziva varijabli. Dok je kod SGPLOT prikaza naziv zavisne varijable paralelan s Y osi, kod GPLOT procedure njen naziv se nalazi u lijevom gornjem kutu. Da bi promijenili prikaz naziva varijable, opet moramo to učiniti u općim postavkama, ovaj put koristeći AXIS naredbu. Razlika između prikaza legendi je nedostatak okvira kod GPLOT procedure, dok je okvir prisutan kod SGPLOT procedure prisutan. Posljednja vidljiva razlika je nedostatak okvira kod GPLOT grafikona. Da bi okvir bio vidljiv moramo koristiti GOPTIONS naredbu u općim postavkama.

Da prikažemo razlike u količini kodiranja između procedura prikazati ćemo linijski grafikon koristeći EUROSTAT podatke. Za prikaz linijskog grafikona koristeći SGPLOT proceduru potrebno nam je bilo napisati 7 linija koda.

```
TITLE 'Razina mlijecne masti i bjelancevina u mlijeku kroz godine';
PROC SGPLOT DATA=dipl.Eurostat;
SERIES X=TIME Y=mm;
SERIES X=TIME Y=prot;
XAXIS LABEL='Godine po mjesecima' GRID;
YAXIS LABEL=' Bjelancevine (%) Mlijecna mast (%)' GRID;
RUN;
```

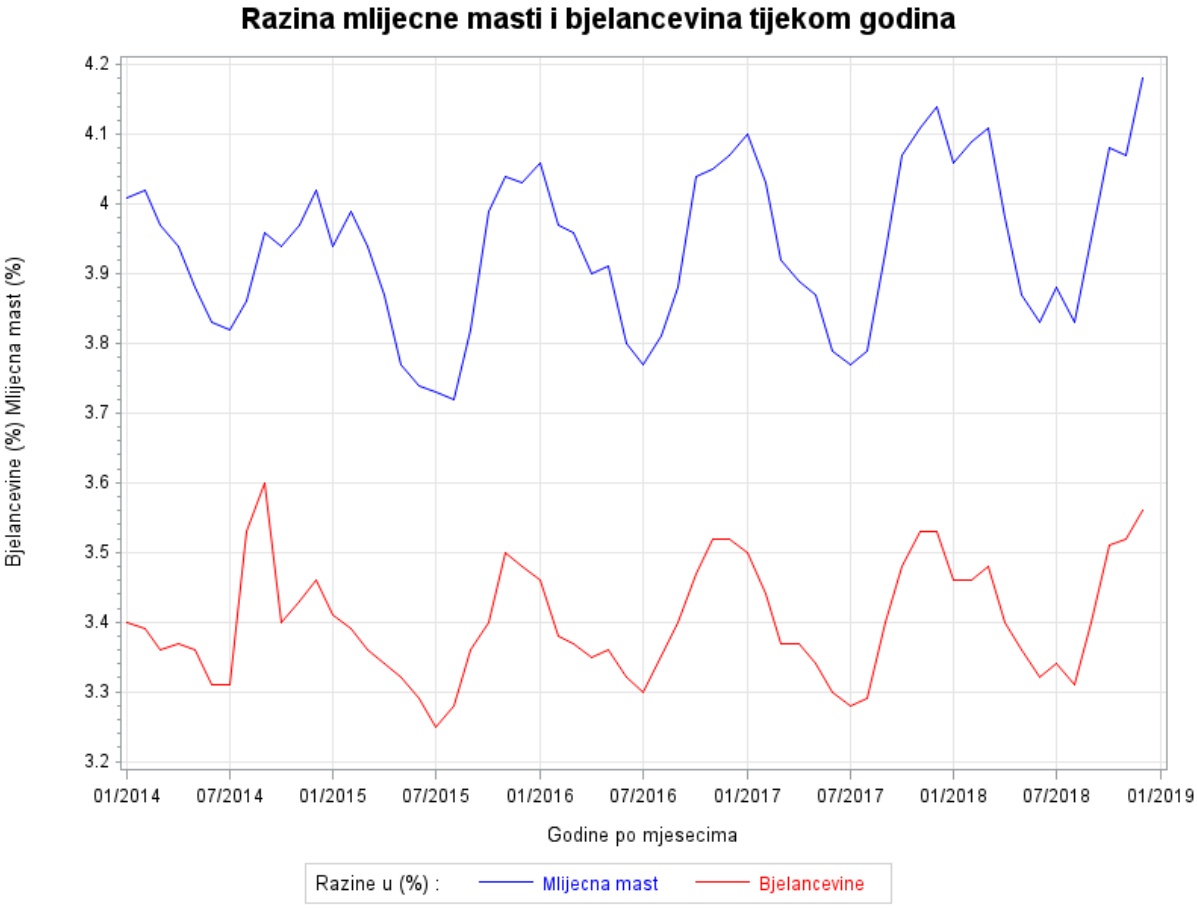


Slika 21. Prikaz linijskog grafikona SGPLOT procedurom

Za usporedbu smo napravili isti grafikon GPLOT procedurom. Dok nam je za SGPLOT proceduru bilo potrebno 7 linija koda, za GPLOT proceduru morali smo napisati 12 linija koda, ali je količina samog kodiranja bila puno veća.

```
AXIS1 LABEL=('Godine po mjesecima' H=1);
AXIS2 LABEL=(A=90 J=C "Bjelancevine (%) Mlijecna mast (%)" H=1);
LEGEND1 POSITION=(OUTSIDE BOTTOM) FRAME
                LABEL= ('Razine u (%) :' H=1)
```

```
                VALUE= (T=1 C=blue 'Mlijecna mast' H=1
                        T=2 C=red 'Bjelancevine' H=1);
GOPTIONS BORDER COLORS=(blue red) INTERPOL=JOIN;
TITLE 'Razina mlijecne masti i bjelancevina tijekom godina';
PROC GPLOT DATA=dipl.Eurostat;
PLOT mm*time prot*time / HAXIS=AXIS1 VAXIS=AXIS2 LEGEND=LEGEND1
OVERLAY GRID;
RUN;
```



Slika 22. Prikaz linijskog višeslojnog grafikona GPLOT procedurom

Glavnu razliku između pisanja koda su dostupne već integrirane opcije unutar SGPLOT procedure to jest manjak takvih opcija unutar GPLOT procedure. Pod time mislimo na opcije definirane unutar općih postavki poput GOPTIONS, AXIS i SYMBOL naredbi, koje je svaki put potrebno resetirati i prilagoditi za podatke koje želimo prikazati GPLOT procedurom (Amruthnath, 2013). U našem primjeru razlika u duljini kodiranja između SGPLOT i GPLOT procedura najviše je uvjetovana upisom AXIS1, AXIS2 i LEGEND opcija unutar općih postavki programa. Za prikaz legende kod SGPLOT procedure nismo morali ništa upisati već

je procedura sama sve odradila za nas. Dok smo kod GPLOT procedure trebali definirati gdje će se nalaziti legenda, što će se nalaziti u njoj, kojim bojama će biti označene vrijednosti i koja će biti veličina slova.

6. Zaključak

U ovom radu možemo zaključiti da nam je dolazak ODS Graphics procedura olakšao i pojednostavio grafički prikaz podataka. Iako još mnogi SAS programeri preferiraju SAS/GRAPH proceduru, ODS Graphics procedure mnogo su bolje za prikaz istog tipa podataka. Samo je pitanje vremena kada će svi programeri preći sa starijih SAS/GRAPH procedura na naprednije SG procedure koje ih u potpunosti mogu zamijeniti.

Glavni problem kod GPLOT procedure je nedostatak i manjak ugrađenih opcija u samu proceduru, što ju čini kompleksnijom za naučiti. Druga stvar zbog čega je GPLOT nezahvalnija procedura je činjenica da za promjene grafičkog prikaza koristimo naredbe u općim postavkama programa. Idući problem je činjenica da prilikom izrade velike količine grafikona moramo svaki put resetirati naredbe u općim postavkama programa kako ne bi došlo do problema prilikom prikaza podataka. Pomoću GPLOT procedure možemo napraviti raspršeni, višeslojni, logaritamski, mjehuričasti i mnoštvo drugih plotova ukoliko dovoljno razumijemo način na koji funkcionira ova procedura.

SGPLOT procedura omogućuje nam izradu već spomenutih, ali i mnoštvo drugih grafikona poput pravokutnih, stupčastih, linijskih, regresijskih, loessa i histograma koje možemo napraviti pomoću ugrađenih opcija u SGPLOT proceduri. Isto tako SGPLOT procedura donosi nam mnoštvo prednosti u odnosu na GPLOT proceduru. Samo neke od prednosti su jednostavnije pisanje koda, bolji prikaz grafikona, više dostupnih opcija grafikona, izrada sofisticiranijih vrsta grafikona, mogućnost preklapanja više od dva grafikona. Osim što se lakše snalaziti po kodu SGPLOT procedure, moguće je puno lakše promijeniti izgled grafikona i podataka koje želimo prikazati. Pod time se misli na činjenicu da se svaka izmjena u pitanju naziva varijabli, boja točaka, izgleda simbola radi unutar procedure, a ne van nje.

Prilikom izrade ovoga rada svakodnevnim radom u SAS programu otkrivali smo nove stvari i shvatili koliko je jednostavno napraviti grafikon koristeći SGPLOT proceduru. Zbog svih navedenih prednosti SGPLOT procedure nad GPLOT procedurom koje smo prikazali i objasnili u ovome radu, možemo zaključiti da je dolaskom ODS Graphics procedura SAS programerima omogućena jednostavnija izrada vizualno privlačnijih grafikona.

7. Literatura

1. Amruthnath S. (2013). PROC SGPLOT over PROC GPLOT. Proceeedings of the Midwest SAS Users Group 2013 Conference. NC: SAS Institute Inc., Paper DV01, 22.-24.09.2013., Columbus, Ohio, str. 1-11.
2. Canchola J., Narra S., Dzidic A. (2018). Customizing and Automating your Graphs using the SAS SG Procedures. Proceedings of the Western Users of SAS software 2018 Conference. Cary, NC: SAS Institute, Inc., Paper 045, 05.-07.09.2018., Sacramento, California, str. 1-20.
3. Cartier J. (2002). The Basics of Creating Graphs with SAS/GRAPH® Software. Proceedings of the SAS Users Group International 27 Conference. Cary, NC: SAS Institute, Inc., Paper 63, 14.-17.04.2002., Orlando, Florida, str. 1-10.
4. Delwiche L. D., Slaughter S. J. (2012). Graphing Made Easy with SG Procedures. Proceedings of the SAS Global Forum 2012 Conference. Cary, NC: SAS Institute, Inc., Paper 259, 22.-25.2012., Orlando, Florida, str. 1-19.
5. Delwiche L. D., Slaughter S. J. (2012). *The Little SAS® Book: A Primer, Fifth Edition*. Cary, NC: SAS Institute Inc.
6. Eurostat (2019). European statistics of the European Union. Luxembourg, Luxembourg. Cow's milk collection and products obtained – monthly data. URL https://ec.europa.eu/eurostat/web/products-datasets/product?code=apro_mk_colm – pristup 10.09.2019.
7. HPA, Godišnje izvješće o uzgoju goveda za 2017. godinu (2017).
8. HPA, Govedarstvo, Izvješće za 2014. godinu (2014).
9. HPA, Govedarstvo, Izvješće za 2015. godinu (2015).
10. HPA, Govedarstvo, Izvješće za 2016. godinu (2016).
11. Kapš M., Lamberson W. R. (Eds.). (2017). *Biostatistics for animal science*. Cabi.
12. Lafler K. P. (2018). Making Your SAS® Results, Reports, Tables, Charts and Spreadsheets More Meaningful with Color. Proceedings of the South-Central SAS Users Group 2018 Conference. Cary, NC: SAS Institute, Inc., 04.-07.11.2018., Austin, Texas, str. 1-20.
13. Larkham I., (2016). SAS Graphing Done Right: Two Good Alternatives to PROC GPLOT/GCHART (Part 2). Proceedings of the Western Users of SAS software 2016 Conference. Cary, NC: SAS Institute, Inc., 07.-09.09.2016., San Francisco, California, str. 1-17.

14. Li L. (2015). One Step Ahead : from PROC GPLOT to PROC SGPLOT. Proceedings of the Golden Horseshoe SAS User Group. Cary, NC: SAS Institute, Inc., 05.2015.
15. Matange S., Heath D. (2011). *Statistical Graphics Procedures by Example: Effective Graphs Using SAS®.* Cary, NC: SAS Institute Inc.
16. Ministarstvo poljoprivrede, Godišnje izvješće o stanju uzgoja goveda u Republici Hrvatskoj za 2018. godinu (2018).
17. SAS Institute Inc. (2016). *SAS/GRAPH® 9.4: Reference, Fifth Edition*. Cary, NC: SAS Institute Inc.
18. Slaughter S. J., Delwiche L. D. (2015). Graphing Made Easy with SGPLOT and SGPANEL Procedures. Proceedings of the SAS Global Forum 2015 Conference. Cary, NC: SAS Institute, Inc., Paper 2441, 26.-29.04.2015., Dallas, Texas, str. 1-19.

Životopis

Sven Papišta rođen je 21.09.1995. godine u Zagrebu. Pohađao je 3. Gimnaziju u Zagrebu od 2010. do 2014. godine. Govori engleski jezik. Upoznat je sa statističkim programima SAS i R u kojima je radio tijekom dva modula na diplomskom studiju. Dugi se niz godina bavio košarkom i nastupao za Košarkaški klub Maksimir.