Časovne vrste

- Kvantitativne metode v geografiji in uvod v GIS -

dr. Gregor Kovačič, doc.

Časovne vrste

- Opazovanje številske spremenljivke v času
- Podatki se nanašajo na zaporedna časovna obdobja
- Statistično vrsto spremenljivke $y_1, y_2, \ldots, y_T$ imenujemo časovna vrsta
- Preučevanje pojavov, ki se časovno spreminjajo
  - Pretok, temperatura, število turistov, plača, brezposelnost …
- Prikaz časovnih vrst je najlažji z linijskimi (črtnimi) grafikoni
  - Ekvidistantna in neakvidistantna časovna vrsta (razmik med podatkoma ni enak)

Časovne vrste so:

- Trenutne – vrednosti se nanašajo na trenutek
  - T zraka ob 21.00 vsak dan
- Intervalne – vrednosti se nanašajo na časovni interval
  - Število diplomantov na leto
- Izvedene – vrednosti so izračunane
  - Povprečni dnevni pretok
- Analiza časovnih vrst:
  - Preučevanje dinamike pojava (časovni razvoj pojavov)
  - Napovedovanje pojava

Indeksi

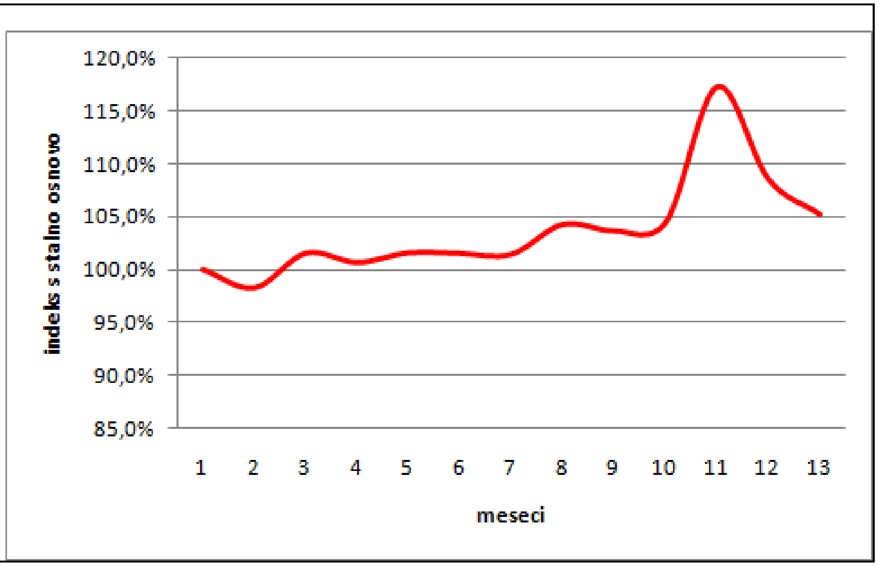
- Osnovno orodje za preučevanje časovnih vrst

1. Indeksi s stalno osnovo
2. Indeksi s premično osnovo
   1. Verižni indeksi
   2. Koeficient rasti in stopnja rasti

Indeks s stalno osnovo *($I_{j/o}$)*

- Primerjava vrednosti časovne vrste (Y_j) z izhodiščno vrednostjo (Y_o)
  - Primer: Primerjava povprečnih mesečnih neto plač v letih 2012 in 2013 z izhodiščno vrednostjo januar 2012
- Najbolje prikazati s črtnim grafikonom
- Indeksi večji od 100 pomenijo višje vrednosti od izhodiščne, manjši od 100 pa manjše

	Bruto plača (v SIT)	Indeks (osnova jan.05)
jan.06	281.593	105,3%
dec.05	290.505	108,6%
nov.05	313.965	117,4%
okt.05	279.506	104,5%
sep.05	277.374	103,7%
avg.05	279.038	104,3%
jul.05	271.419	101,4%
jun.05	271.654	101,5%
maj.05	271.814	101,6%
apr.05	269.368	100,7%
mar.05	271.717	101,6%
feb.05	262.911	98,3%
jan.05 - Y_0	267.544	100,0%



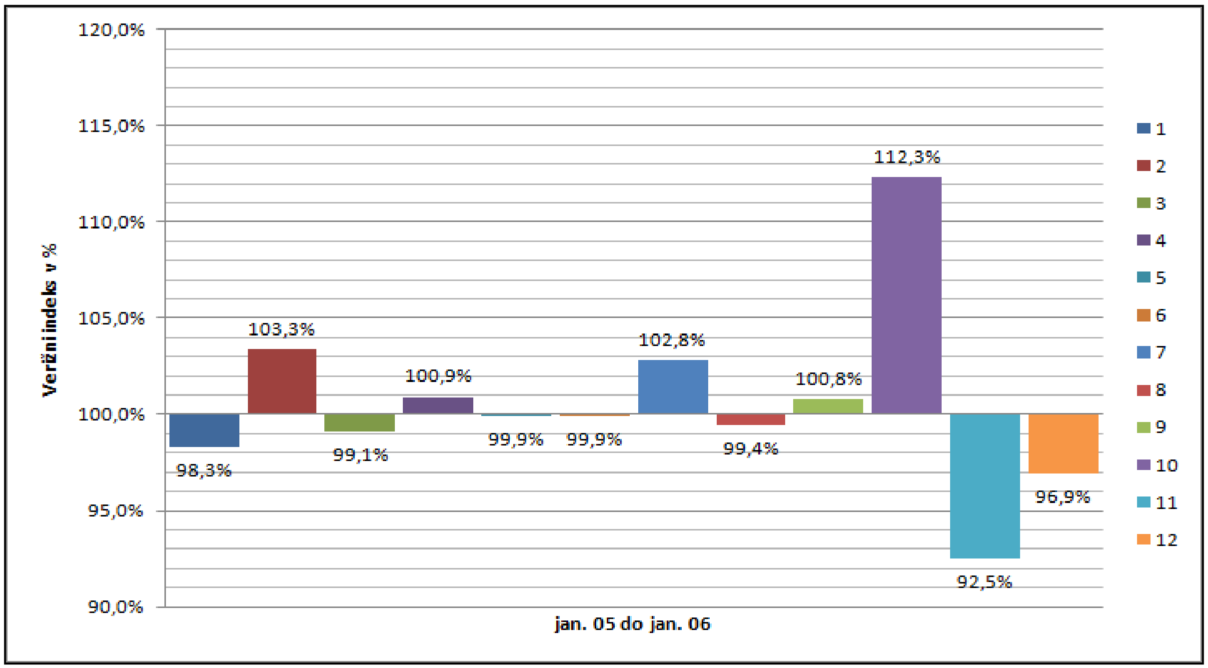
$$I_{j/o} = 100 \times \frac{Y_j}{Y_o}$$

Verižni indeks (V_j)

- Primerjava vrednosti časovne vrste (Y_j) z vrednostjo v predhodnem časovnem obdobju (Y_{j-1})
  - Primer: Primerjava povprečnih mesečnih neto plač v letih 2012 in 2013
- Najbolje prikazati s stolpčnim grafikonom
- Indeksi večji od 100 pomenijo povečanje v primerjavi s predhodnim časovnim obdobjem, vrednosti manjše od 100 pa zmanjšanje v primerjavi s predhodnim časovnim obdobjem

$$V_j = 100 \times \frac{Y_j}{Y_{j-1}}$$

	Bruto plača (v SIT)	Verižni indeks
jan.06	281.593	96,9%
dec.05	290.505	92,5%
nov.05	313.965	112,3%
okt.05	279.506	100,8%
sep.05	277.374	99,4%
avg.05	279.038	102,8%
jul.05	271.419	99,9%
jun.05	271.654	99,9%
maj.05	271.814	100,9%
apr.05 - Y_j	269.368	99,1%
mar.05 - Y_{j-1}	271.717	103,3%
feb.05	262.911	98,3%
jan.05	267.544	/



Koeficient rasti (K_j) in stopnja rasti (S_j)

Koeficient rasti (K_j) dobimo, če primerjamo dva zaporedna podatka časovne vrste (Y_j in Y_{j-1})

- $K_j = V_j : 100$ – izpeljava

Stopnja rasti (S_j)

- Izraža spremembo med dvema zaporednima podatkoma časovne vrste (Y_j in Y_{j-1})
- Je razlika med dvema zaporednima podatkoma časovne vrste, deljeno s predhodnim podatkom ter pomnoženo s 100, da dobimo izraženo v %
- $S_j = V_j - 100$ ali $(K_j - 1) \times 100$ – izpeljavi
- Povprečna stopnja rasti je geometrijska sredina verižnih indeksov, zmanjšana za 100 ali = - 100
  - Vrednosti manjše od 0 kažejo na upad, večje na rast

	Bruto plača (v SIT)	Indeks (osnova jan.05)	Verižni indeks	Stopnja rasti	Koeficient rasti
jan.06	281.593	105,3%	96,9%	-3,1	0,969
dec.05	290.505	108,6%	92,5%	-7,5	0,925
nov.05	313.965	117,4%	112,3%	12,3	1,123
okt.05	279.506	104,5%	100,8%	0,8	1,008
sep.05	277.374	103,7%	99,4%	-0,6	0,994
avg.05	279.038	104,3%	102,8%	2,8	1,028
jul.05	271.419	101,4%	99,9%	-0,1	0,999
jun.05	271.654	101,5%	99,9%	-0,1	0,999
maj.05	271.814	101,6%	100,9%	0,9	1,009
apr.05	269.368	100,7%	99,1%	-0,9	0,991
mar.05	271.717	101,6%	103,3%	3,3	1,033
feb.05	262.911	98,3%	98,3%	-1,7	0,983
jan.05	267.544	100,0%	/	/	/

$$K_j = \frac{Y_j}{Y_{j-1}}$$

$$S_j = \frac{Y_j - Y_{j-1}}{Y_{j-1}} \times 100$$

$$\bar{S} = \bar{V} - 100$$

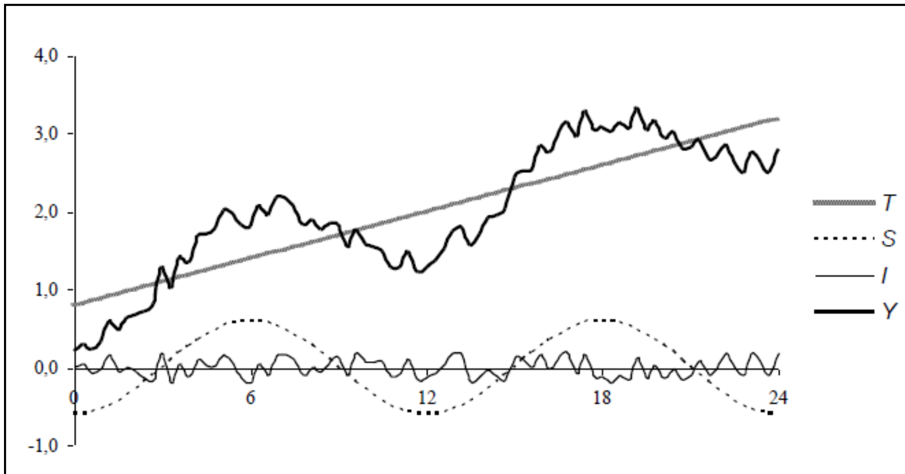
$$\bar{V} = \sqrt[M]{V_1 \times V_2 \times \ldots \times V_M}$$

Komponente časovne vrste

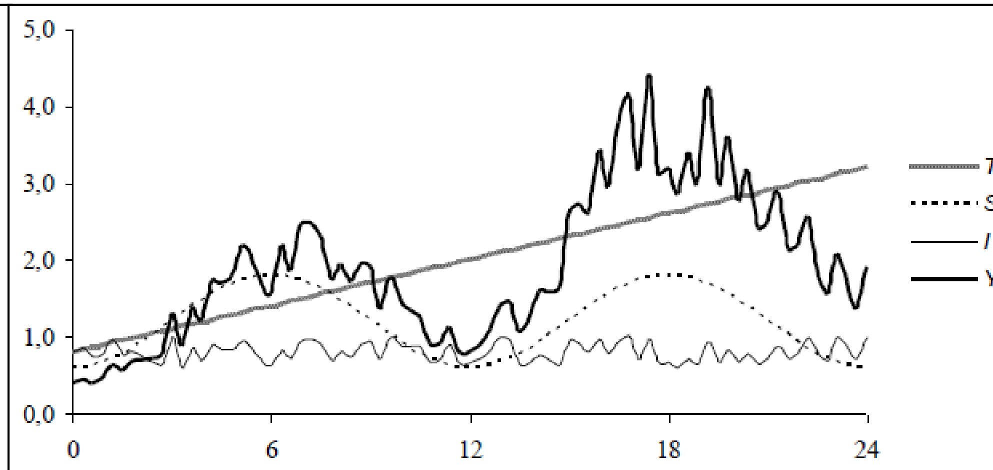
- Trend (*T*)
  - Kaže dolgoročno gibanje pojava
  - Glej naprej!
- Sezonska komponenta (*S*)
  - Se nanaša na ponavljajoče se (periodične) spremembe zaradi letnih časov, praznikov, vremena, dnevnega gibanja prejete količine sončevega obsevanja …
  - Ima svojo dolžino, dolžina sezonske komponente je konstantna
  - Je predvidljiva. Navadno eno koledarsko leto.
- Ciklična komponenta (*C*)
  - Neperiodične spremembe, gibanja okrog trenda
  - Dolžina cikla ni določena, a je daljša od enega leta oz. daljših časovnih obdobjih
  - Primer: Spremembe v gospodarstvu, okolju, modi …
- Slučajna komponenta (*I*) – neregularni vplivi, šum
  - Neznani dejavniki, ki jih ne moremo razložiti s trendom, sezonsko in/ali ciklično komponento

Komponente časovne vrste

- Časovna vrsta ni nujno rezultat delovanja vseh komponent
  - Slučajna komponenta je vedno prisotna
  - Skoraj vedno sta prisotna trend in sezonskost
- Teorija časovnih vrst skuša identificirati komponente, jih razumeti in uporabiti
- Poznamo več modelov časovnih vrst
  - Aditivni: $Y = T + S + C + I$
  - Multiplikativni: $Y = T \times S \times C \times I$ oz. $\log Y = \log T \times \log S \times \log C \times \log I$



Slika 32: Aditivni model $Y = T + S + I$ za časovno vrsto



Slika 33: Multiplikativni model $Y = T \cdot S \cdot I$ za časovno vrsto

Metoda drsečih sredin

- Med tehnike glajenja podatkov štejemo postopke, s katerimi poskušamo izločiti vpliv slučajne komponente, da bi odkrili trend, sezonsko in ciklično komponento
  - Metoda drsečih sredin
  - Eksponentno glajenje (samo omenjamo, ne bomo računali)

Metoda drsečih sredin

- Originalni časovni vrsti $y_1, y_2, \ldots, y_T$ prilagodimo časovno vrsto drsečih sredin, katere členi so zaporedja povprečij, izračunanih iz dela niza podatkov opazovane časovne vrste.
- V grafu na novo kreirane časovne vrste se vidi, kako se zgladijo nihanja v podatkih in jasneje pokaže trend
- Najprej določimo red drsečih sredin (*r*)
  - Vsaka drseča sredina je na sredini intervala dolžine *r*

Način izračunavanja drsečih sredin
Oznake:
r – red drsečih sredin
Y^r, npr. drseče sredine reda 3 označimo Y^3

r je liho število, $r = 2i + 1$

$$y_k^r = \frac{1}{r}(y_{k-i} + y_{k-i+1} + \ldots + y_k + \ldots + y_{k+i-1} + y_{k+i})$$

b) r je sodo število, $r = 2i$

$$y_k^r = \frac{1}{r}(\frac{1}{2}y_{k-i} + y_{k-i+1} + \ldots + y_k + \ldots + y_{k+i-1} + \frac{1}{2}y_{k+i})$$

$$Y^r = y_{i+1}^r, y_{i+1}^r, \ldots, y_{T-i}^r$$

Metoda drsečih sredin

- Originalni časovni vrsti prilagodimo časovno vrsto drsečih sredin, katere členi so zaporedja povprečij, izračunanih iz dela niza podatkov opazovane časovne vrste.
- V grafu na novo kreirane časovne vrste se vidi, kako se zgladijo nihanja v podatkih in jasneje pokaže trend
- Najprej določimo red drsečih sredin (*r*)
  - Vsaka drseča sredina je na sredini intervala dolžine r

$$r = 3 = 2 \cdot 1 + 1 \qquad (i{=}1)$$
$$y_2^3 = \frac{1}{3}(830 + 1026 + 853) = 903{,}0$$
$$y_3^3 = \frac{1}{3}(1026 + 853 + 651) = 843{,}3$$
...
$$r = 4 = 2 \cdot 2 \qquad (i{=}2)$$
$$y_3^4 = \frac{1}{4}(\frac{1}{2} \cdot 830 + 1026 + 853 + 651 + \frac{1}{2} \cdot 545) = 804{,}375$$
$$y_4^4 = \frac{1}{4}(\frac{1}{2} \cdot 1026 + 853 + 651 + 545 + \frac{1}{2} \cdot 818) = 742{,}750$$
...
$$r = 6 = 2 \cdot 3 \qquad (i{=}3)$$
$$y_4^6 = \frac{1}{6}(\frac{1}{2} \cdot 830 + 1026 + 853 + 651 + 545 + 818 + \frac{1}{2} \cdot 1089) = 808{,}750$$
$$y_5^6 = \frac{1}{6}(\frac{1}{2} \cdot 1026 + 853 + 651 + 545 + 818 + 1089 + \frac{1}{2} \cdot 849) = 815{,}583$$

Primer: prirast smreke po letih, dolžina časovne vrste =15

Leto	Prirast (mm^2)	Drseče sredine r=3	Drseče sredine r=4	Drseče sredine r=6
1	830			
2	1026	903,0		
3	853	843,3	804,4	
4	651	683,0	742,8	808,8
5	545	671,3	746,3	815,6
6	818	817,3	800,5	810,4
7	1089	918,7	878,1	794,0
8	849	968,7	871,1	791,0
9	968	718,7	777,8	824,3
10	339	709,3	755,8	844,3
11	821	700,3	797,1	857,3
12	941	989,7	895,4	894,8
13	1207	1011,3	1033,8	
14	886	1158,0		
15	1381			

r je liho število, $r = 2i + 1$

$$y_k^r = \frac{1}{r}(y_{k-i} + y_{k-i+1} + \ldots + y_k + \ldots + y_{k+i-1} + y_{k+i})$$

b) r je sodo število, $r = 2i$

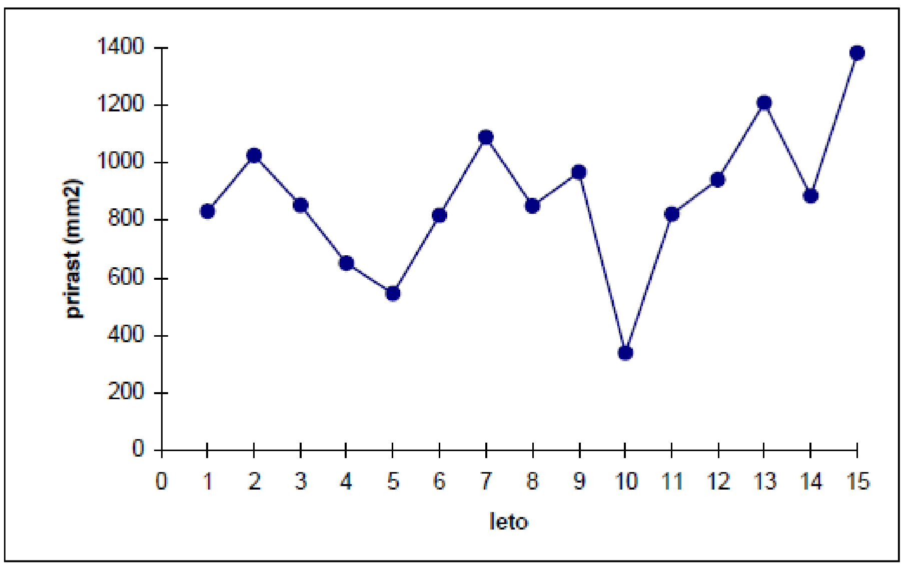
$$y_k^r = \frac{1}{r}(\frac{1}{2} y_{k-i} + y_{k-i+1} + \ldots + y_k + \ldots + y_{k+i-1} + \frac{1}{2} y_{k+i})$$
$$Y^r = y_{i+1}^r, y_{i+1}^r, \ldots, y_{T-i}^r$$

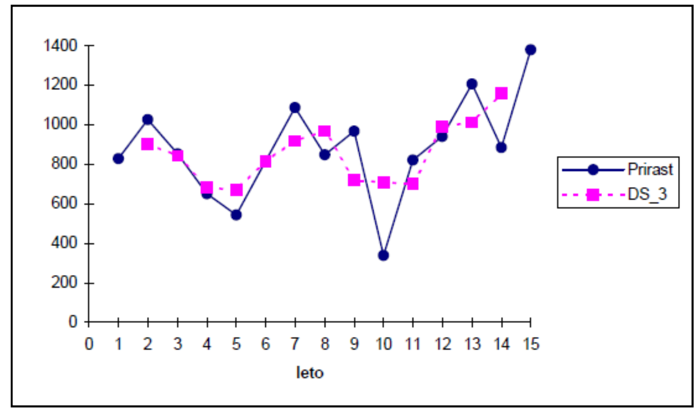
Metoda drsečih sredin

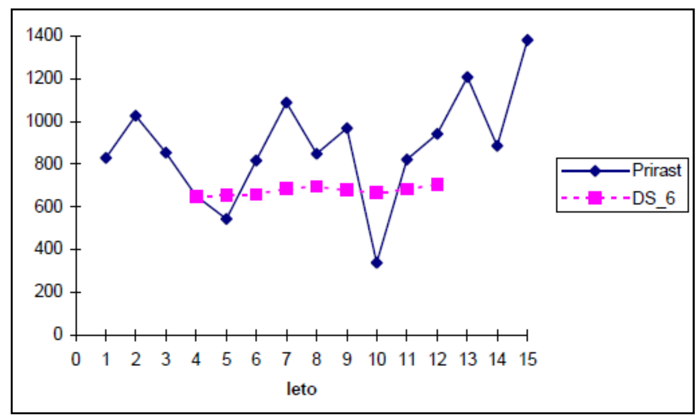
- Večji red *r* povzroči večje glajenje osnovne časovne vrste
- Časovna vrsta drsečih sredin je krajša od osnovne časovne vrste
- Prednost je ta, da je mogoče odkrivati trend
  - Če je trend osnovne ČV linearen, ČV drsečih sredin prikazuje trend
  - Če ima ČV periodično oz. ciklično komponento, ČV drsečih sredin prikazuje trend, če je njen red *r* enak dolžini periode

Primer: prirast smreke po letih, dolžina časovne vrste =15

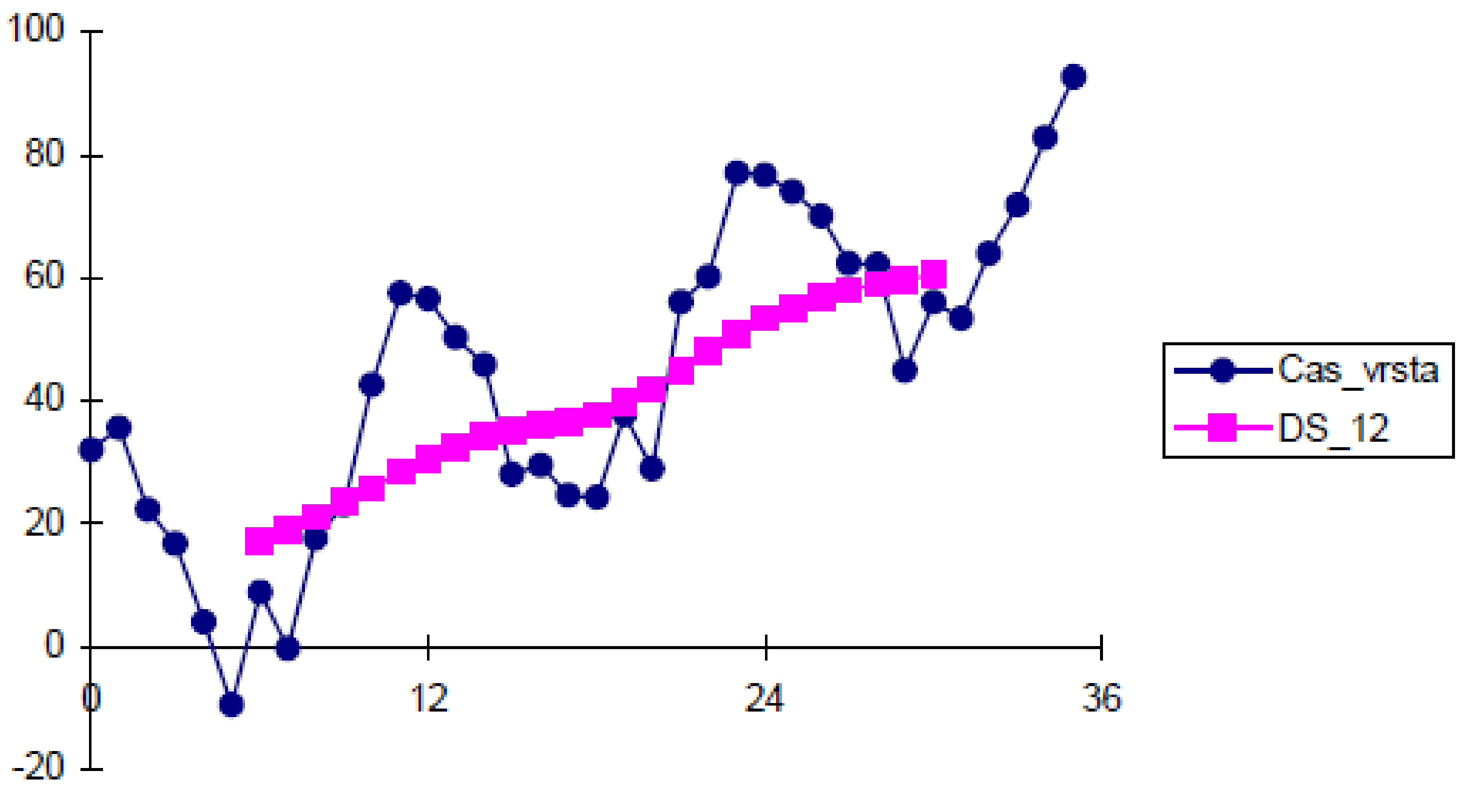
Leto	Prirast (mm^2)	Drseče sredine r=3	Drseče sredine r=4	Drseče sredine r=6
1	830			
2	1026	903,0		
3	853	843,3	804,4	
4	651	683,0	742,8	808,8
5	545	671,3	746,3	815,6
6	818	817,3	800,5	810,4
7	1089	918,7	878,1	794,0
8	849	968,7	871,1	791,0
9	968	718,7	777,8	824,3
10	339	709,3	755,8	844,3
11	821	700,3	797,1	857,3
12	941	989,7	895,4	894,8
13	1207	1011,3	1033,8	
14	886	1158,0		
15	1381			







Metoda drsečih sredin in trend

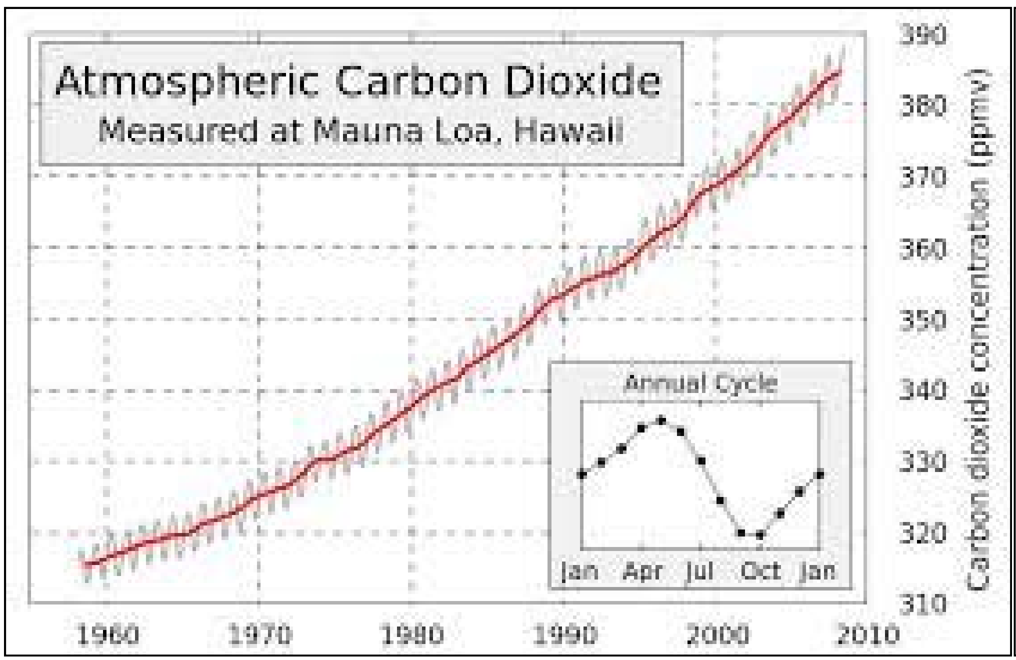


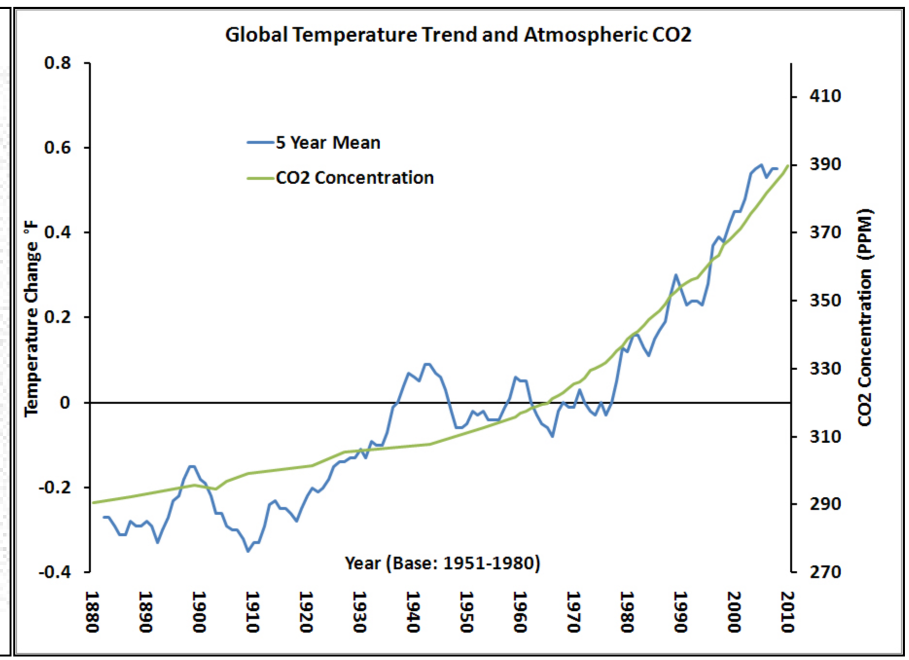
Slika: Časovna vrsta s periodo 12 in drseče sredine reda 12

Če je red (*r*) časovne vrste drsečih sredin enak dolžini periode, le ta prikazuje trend. Ciklična komponenta se izniči.

Trend (*T*)

- Kaže dolgoročno gibanje pojava
- Podaja smer (premica ali krivulja) razvoja
- Trend je lahko naraščajoč, padajoč ali stacionaren
- Dolgoročne spremembe so lahko pogojene s spremembo podnebja, okolja, gospodarskih sprememb …





Metode za določanje trenda (*T*)

- Prostoročno določanje trenda na grafičnem prikazu
  - Dobro za začetek analize
- Z metodo glajenja (drseče sredine)
  - Zgolj v določenih primerih (periode enake redu drsečih sredin (*r*))
- Analitično z metodami regresijske analize
  - Metoda najmanjših kvadratov
  - Linearni in nelinearni regresijski model (MS Excel pozna več možnosti) za določanje trenda

Ugotavljanje trenda z metodo najmanjših kvadratov

$$\boxed{\sum_{t=1}^{N}(Y_t - T_t)^2 = min}$$

Y - opazovani pojav
t - čas

- Uvedemo transformacijo časovne vrste tako, da ordinatno os premaknemo na njeno sredino
  - Uvedemo t. i. tehnični čas, zanj velja $\sum_{t=1}^{N} x_t = 0$

Način določanja intervalov tehničnega časa za liho in sodo število letnih obdobij:

	Leto	2000	2001	2002	2003	2004	2005	2006	2007	$\sum x_t$
$N=2k+1$	x_t		–3	–2	–1	0	1	2	3	0
$N=2k$	x_t	–3,5	–2,5	–1,5	–0,5	0,5	1,5	2,5	3,5	0

Linearni regresijski model za določanje trenda

- $Y = a + bt + \varepsilon$
- Parameter *a* je aritmetična sredina pojava v opazovanem obdobju
- Parameter *b* pove, za koliko se spremeni pojav v povprečju, če se čas spremeni za eno enoto

$$a = \frac{\sum_{t=1}^{N} Y_t}{N} = \overline{Y} = M_Y$$

$$b = \frac{\sum_{t=1}^{N} Y_t \cdot x_t}{\sum_{t=1}^{N} x_t^2}$$

Predhodna predpostavka: Iz grafa ugotovimo (lahko s pomočjo grafa v MS Excelu), da je linearni trend sprejemljiv.

Linearni regresijski model za določanje trenda - primer

- Dolžina časovne vrste je liho število

b = 2237 / 60 = 37,28

a = 8108 / 9 = 900,88

T_t = 901 + 37,28 • x_t

- Povprečno število prepeljanih potnikov v obravnavanem obdobju je bilo 901.000 (*a*)
- V povprečju se je v obravnavanem obdobju število prepeljanih potnikov vsako leto povečalo za 37.280 (*b*).

x_t – tehnični čas

T_t – premica linearnega trenda

Y_t – opazovan pojav v časovni enoti

$$a = \frac{\sum_{t=1}^{N} Y_t}{N} = \overline{Y} = M_Y$$

$$b = \frac{\sum_{t=1}^{N} Y_t \cdot x_t}{\sum_{t=1}^{N} x_t^2}$$

Število z letali prepeljanih potnikov v R Sloveniji v letih od 1999 do 2007 ter izračun trenda

Leto	**Število potnikov** (v 1000)	x_t	$Y_t \cdot x_t$	x_t^2	T_t
1999	780	–4	–3.120	16	752
2000	866	–3	–2.598	9	789
2001	801	–2	–1.602	4	826
2002	814	–1	–814	1	864
2003	864	0	0	0	901
2004	885	1	885	1	938
2005	944	2	1.888	4	975
2006	1.018	3	3.054	9	1.013
2007	1.136	4	4.544	16	1.050

Linearni regresijski model za določanje trenda - primer

<u>Izračun napovedi za leto 2012</u>

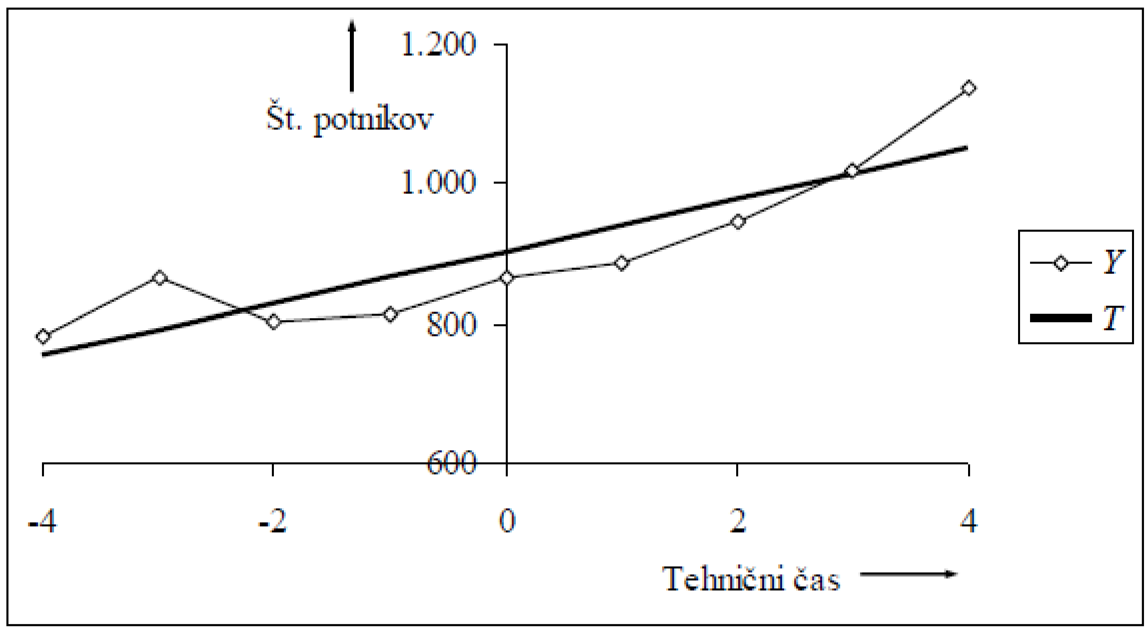
$T_t = 901 + 37{,}28 \bullet x_t$

- $T_{2012} = T_9 \rightarrow$

$901 + 37{,}28 \cdot \boxed{9} = 1236{,}5$

- Napoved kaže, da je bilo leta 2012 prepeljanih 1.236.500 potnikov

Leto	Število potnikov (v 1000)	x_t	$Y_t \cdot x_t$	x_t^2	T_t
1999	780	–4	–3.120	16	752
2000	866	–3	–2.598	9	789
2001	801	–2	–1.602	4	826
2002	814	–1	–814	1	864
2003	864	0	0	0	901
2004	885	1	885	1	938
2005	944	2	1.888	4	975
2006	1.018	3	3.054	9	1.013
2007	1.136	4	4.544	16	1.050



T – premica linearnega trenda
Y – opazovan pojav v časovni enoti

Ocenjevanje kakovosti trenda

$$\sigma_T = \sqrt{\sigma^2} = \sqrt{\frac{\sum (Y_t - T_t)^2}{N}}$$

$$KV_T = \frac{\sigma_T}{\mu_Y} \text{ x } 100$$

<u>Ocenjujemo z dvema merama</u>

- Standardnim odklonom trenda (σ_T) in
- Koeficientom variacije trenda (KV_T)

Manjše vrednosti obeh mer odražajo ustreznejšo funkcijo trenda. Pozor - mejne vrednosti ne obstojijo, zato uporabno predvsem za primerjavo različnih trendnih funkcij (linearna, nelinearna …) znotraj iste časovne vrste.

<u>Izračun za primer ocene kakovosti linearnega trenda</u>

- $T_t = 901 + 37{,}28 \bullet x_t$
- $\sigma = 49.910$
- KV (%) = 5,54 %

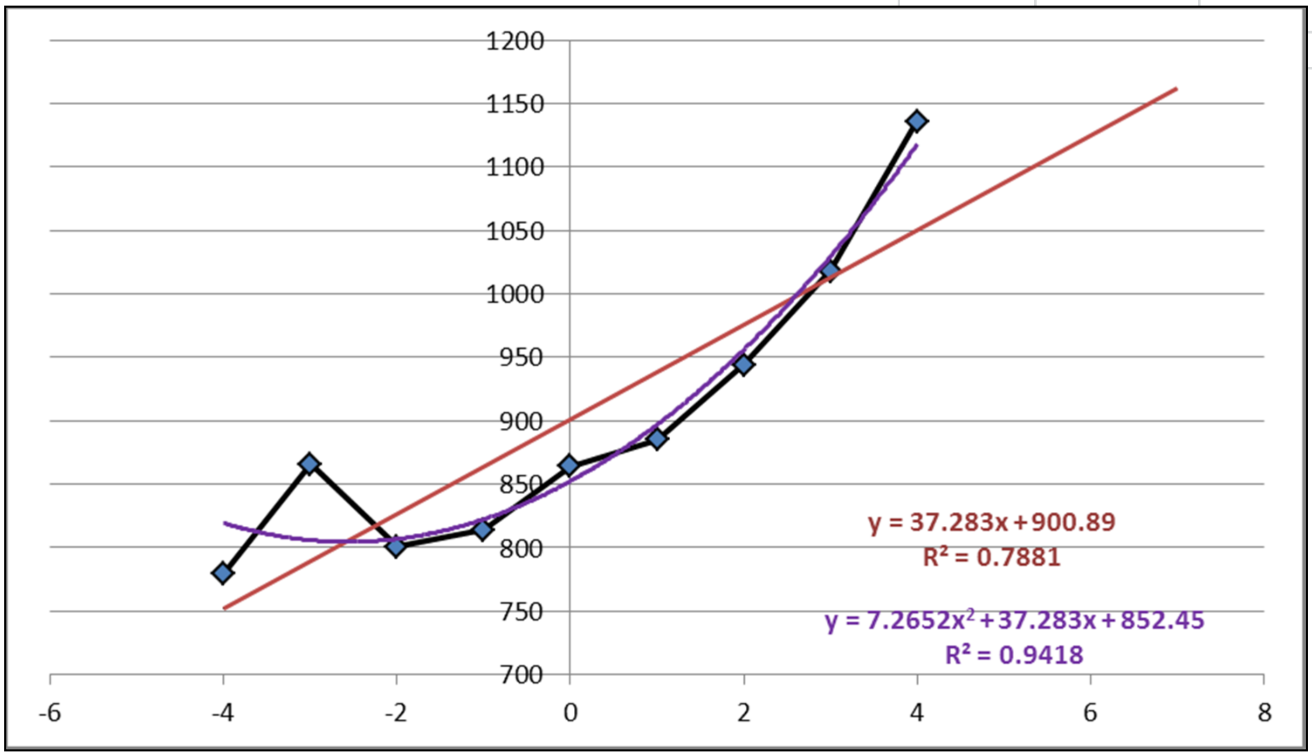
Leto	Število potnikov - Y_t (v 1000)	x_t	$Y_t \bullet x_t$	x_t^2	T_t	$(Y_t - T_t)^2$
1999	780	-4	-3120	16	752	797.61
2000	866	-3	-2598	9	789	5922.69
2001	801	-2	-1602	4	826	641.30
2002	814	-1	-814	1	864	2460.85
2003	864	0	0	0	901	1360.87
2004	885	1	885	1	938	2827.37
2005	944	2	1888	4	975	989.48
2006	1018	3	3054	9	1013	27.68
2007	1136	4	4544	16	1050	7392.22
Vsota	8108	0	2237	60	8108	22420
					σ =	49.91
					KV (%) =	5.54

Ocenjevanje kakovosti trenda

Izračun za primer ocene kakovosti paraboličnega trenda

- T_t = 852,45 + 37,283 • x_t+ 7,2652 • x_t^2
- σ = 26.170
- *KV* (%) = 2,90 %

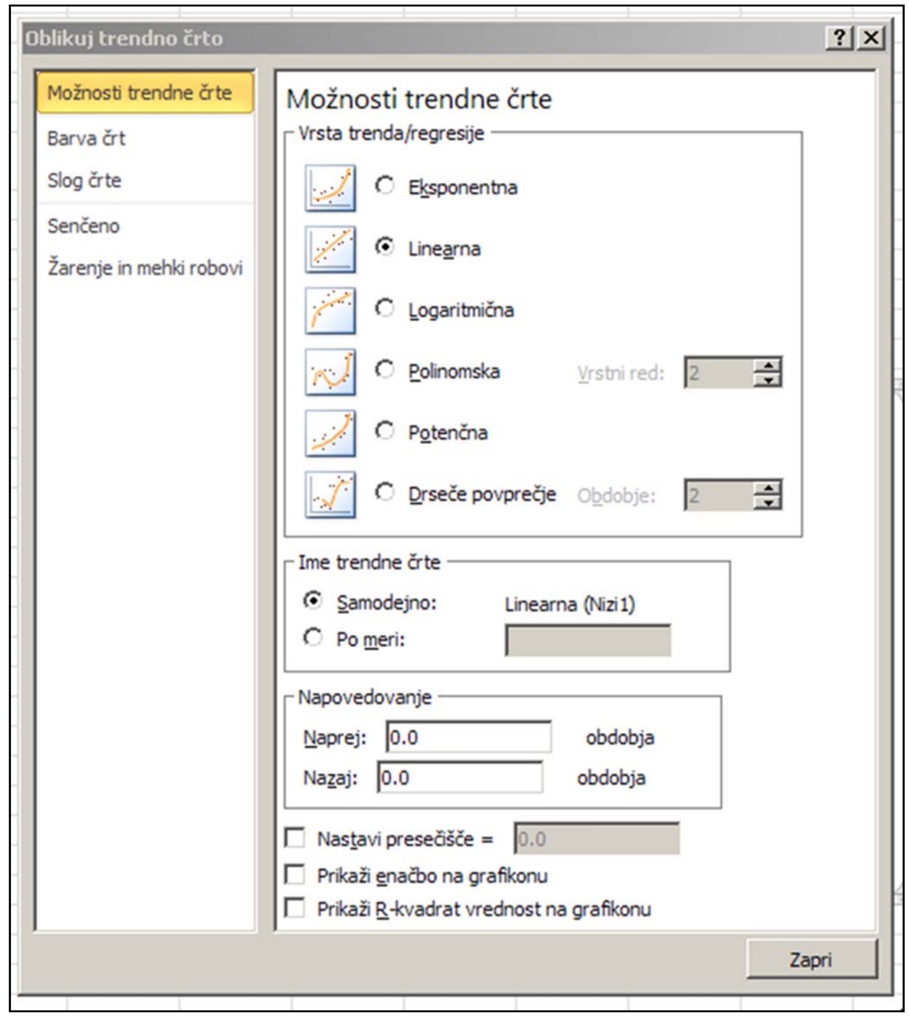
Leto	Število potnikov - Y_t (v 1000)	x_t	$Y_t \cdot x_t$	x_t^2	T_t	$(Yt - Tt)^2$
1999	780	-4	-3120	16	820	1565.09
2000	866	-3	-2598	9	806	3601.46
2001	801	-2	-1602	4	807	35.34
2002	814	-1	-814	1	822	71.10
2003	864	0	0	0	852	133.40
2004	885	1	885	1	897	143.96
2005	944	2	1888	4	956	145.85
2006	1018	3	3054	9	1030	136.56
2007	1136	4	4544	16	1118	330.32
Vsota	8108	0	2237	60	8108	6163
					σ =	26.17
					KV (%) =	2.90



- Vrednost determinacijskega koeficienta za parabolično funkcijo je višja (r^2 = 0,94) → boljše prileganje in natančnejše napovedovanje
- Vrednosti standardnega odklona trenda in KV trenda sta pri parabolični funkciji nižji → ustreznost trendne premice (funkcije trenda) je tako višja

Določanje trenda s pomočjo MS Excela

- S pomočjo MS Excela lahko naredimo izbor prave vrste trendne črte z namenom, da izboljšamo kakovost funkcije trenda (boljše prileganje in natančnejše napovedovanje)
- MS Excel pozna naslednje trendne črte: linearno, logaritmično, polinomsko (od 2. do 6. stopnje), eksponentno in potenčno
- Možnost prikaza funkcije (enačbe) ter izračun determinacijskega koeficienta → možnost prezkušanja in primerjanja na poti k izbiri optimalnega rezultata
- Prikaz trendne črte poljubno za naprej in nazaj
- Funkcija TREND(known_y‘s;known_x‘s;new_x‘s;const) vrne vrednost vzdolž linearnega trenda



Trendne črte (funkcije) v MS Excelu

- **Linearna trendna črta** je ravna črta z najboljšim ujemanjem, ki se uporablja za preproste linearne podatkovne serije. Podatki so linearni, če je vzorec v podatkovnih točkah podoben črti. Linearna trendna črta navadno prikazuje, da nekaj enakomerno narašča ali pada.
- **Logaritmična trendna črta** je krivulja z najboljšim ujemanjem, ki je najbolj uporabna pri hitri stopnji spreminjanja z višanjem ali padanjem podatkov in nato popolnem prenehanju rasti. Logaritmična trendna črta lahko uporabi tako negativne kot tudi pozitivne vrednosti.
- **Polinomska trendna črta** je krivulja, ki se uporablja pri nestalnih podatkih. Uporabna je na primer za analizo dobičkov in izgub pri velikem nizu podatkov. Vrstni red polinoma lahko določa število nestalnosti podatkov ali pa število ovinkov (hribi in doline) na krivulji. Vrstni red 2 polinomske trendne črte ima običajno samo en hrib ali dolino. Vrstni red 3 pa običajno dva hriba ali dve dolini. Vrstni red 4 običajno največ tri hribe in doline.
- **Eksponentna trendna črta** je krivulja, ki je najbolj uporabna, ko vrednosti podatkov naraščajo ali padajo s konstantno naraščajočo stopnjo. Če podatki vsebujejo ničelne ali negativne vrednosti, eksponentne trendne črte ni mogoče ustvariti.
- **Trendna črta moči (potence)** je krivulja, ki se najbolje uporablja za nize podatkov za primerjavo mer, ki naraščajo pri določeni stopnji na primer pospešek dirkalnega avtomobila pri sekundnih intervalih. Če podatki vsebujejo ničelne ali negativne vrednosti, trendne črte moči ni mogoče ustvariti.

Literatura in viri

- Ferligoj, Anuška. 1995. *Osnove statistike na prosojnicah.* Ljubljana: Samozaložba Z. Batagelj.
- Statistika. 2013. »*Electronic statistic textbook, StatSoft*«. Http://www.statsoft.com.
- Rogerson, Peter A. 2006. *Statistical Methods for Geography: a student guide*. London: Sage Publications.
- Kastelec, Damijana in Katarina Košmelj. 2010. *Osnove statistike z Excelom 2007.* Ljubljana: Biotehniška fakulteta. - Dostopno tudi na medmrežju.
- Andoljšek, Žiga. 2004. *Statistika: obrazci in postopki.* Ljubljana: Fakulteta za upravo Univerze v Ljubljani.
- Gams, Anton in Marko Gams. 2009. *Poslovna informatika s statistiko.* Ljubljana: Zavod IRC. Http://www.dlib.si/details/URN:NBN:SI:doc-64ERKYRT.
- Napredne učne kocke. 2014. Http://www.nauk.si.