FAKULTETA ZA ORGANIZACIJSKE VEDE KRANJ

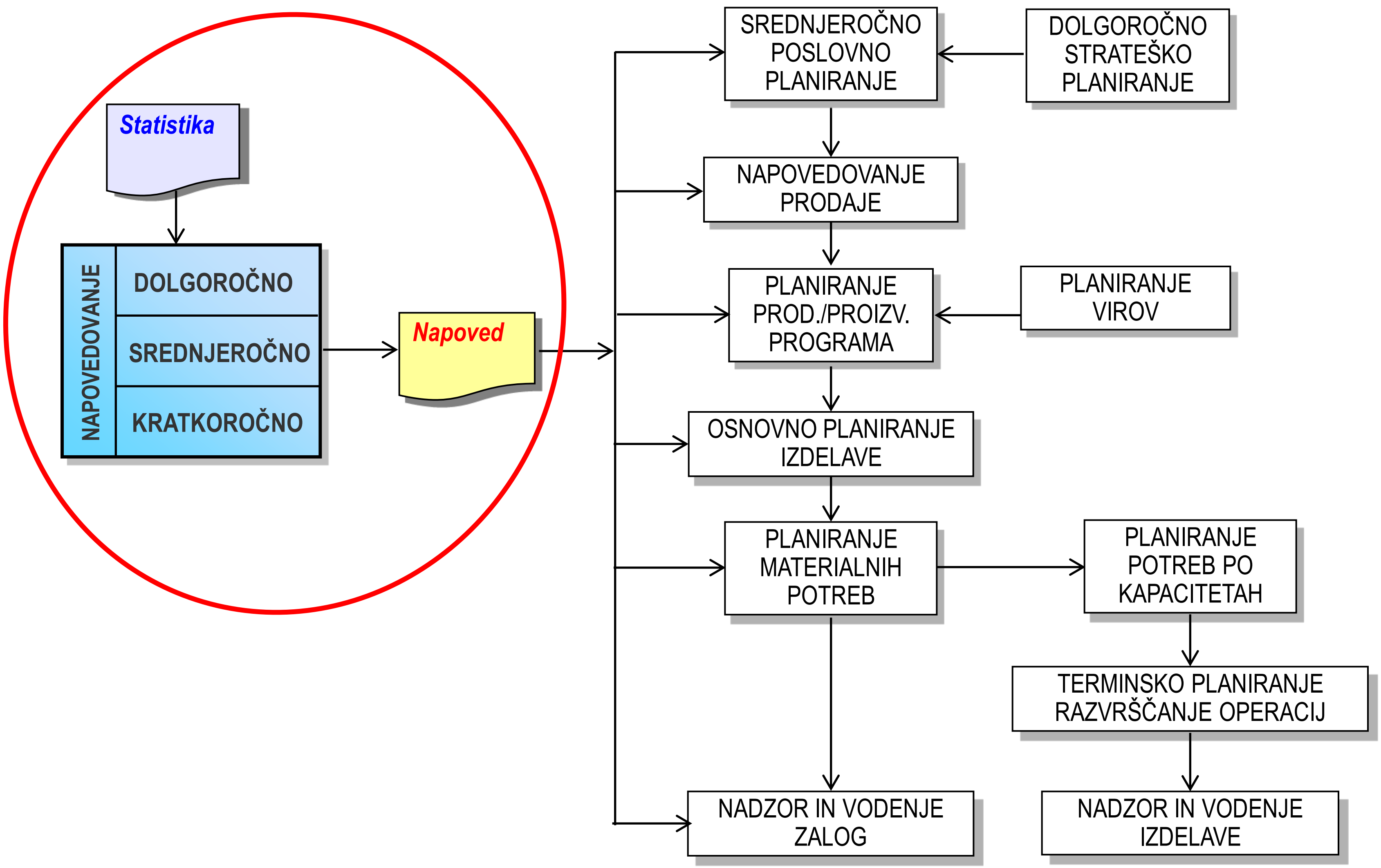
Katedra za poslovne in delovne sisteme

Matjaž ROBLEK

METODE IN TEHNIKE PLANIRANJA

03 Napovedovanje – stohastično planiranje

Mesto napovedovanja v sistemu planiranja in vodenja



Napovedovanje

Forecasting

- je ocenjevanje, predvidevanje (anticipacija, predikcija) mogočih bodočih dogodkov oziroma dejavnosti.
- ker imajo poslovni procesi vztrajnost, se sme iz dogajanj v preteklosti preko trenutnih dogajanj sklepati na dogajanja v prihodnosti.
- mnogo poslovnih procesov ima stohastični značaj, spreminjajo se po kraju in času v skladu z zakoni verjetnosti, zato se napovedi dogajanj v prihodnosti podrejajo stohastičnim zakonitostim in niso popolnoma zanesljive,
- zato napovedovanje imenujemo tudi stohastično planiranje.

Napovedovanje se uporablja za

- napovedovanje povpraševanja oziroma prodaje pri planiranju prodajno/proizvodnega programa in pri osnovnem planiranju izdelave,
- napovedovanje porabe materialov (ali sestavnih delov) pri planiranja materialnih potreb,
- avtomatsko popolnjevanje zalog materialnih postavk razreda C / X,
- napovedovanje trajanja izdelavnega ciklusa za izdelke/storitve in dobavnega časa za materiale,
- napovedovanje slabe kakovosti (izmeta),
- napovedovanje razpoložljivosti izdelavnih zmogljivosti (napoved izpada delovnih sredstev zaradi okvar, odsotnosti delavcev) ipd.

Metode napovedovanja glede na planski horizont

- tehnološko napovedovanje, ki sega nekaj desetletij v prihodnost (vezano na politiko),
  - dolgoročno predvideva znanstvene in inovacijske možnosti, brez razmišljanja o aplikaciji ugotovitev (npr. metoda Delphi);
- dolgoročno napovedovanje za obdobje 3 do 10 let v naprej (vezano na strategijo),
  - ocenjuje razvoj poslovnih procesov ob predpostavki, da je ocenjeno prevladujoče gibalo znanstvenih in inovacijskih rešitev (niso izključeni tudi drugi motivi);
- kratkoročno napovedovanje s planskim obdobjem enega do dveh let ali manj (uporabno pri taktiki in operativi planiranja),
  - skuša predvideti **dogodke** , ki lahko vplivajo na nek poslovni procesa v naslednjem kratkoročnem obdobju.

Oblike <u>dogodkov</u> (pojavov)

- Dogodki so lahko:
  - <u>endogeni</u>, če so odvisni izključno od časa;
    - ker je čas spremenljivka, na katero ni mogoče vplivati, a je predvidljiva, so pojavi <u>stacionarni</u> (razmeroma stabilni) in jih je mogoče <u>napovedovati dokaj zanesljivo</u>, (npr. povpraševanje (prodaja) po osnovnih živilih)
  - <u>eksogeni</u>, če nanje poleg časa vplivajo še drugi bolj ali manj nepoznani dejavniki, predvsem ekološkega in sociološkega značaja;
    - ker so ti dejavniki nepredvidljivi, so taki pojavi nestabilni (<u>nestacionarni</u>) in jih ni mogoče zanesljivo napovedovati, (npr. povpraševanje po modnih artiklih)

Metode kratkoročnega napovedovanja glede na oblike dogodkov (pojavov)

- Kvalitativne metode so neformalne in temeljijo na izkušnjah in subjektivnih ocenah;
  - izkustveno (heuristično) ocenjevanje, kadar o pojavu, ki se želi napovedovati, ni na razpolago nobenih eksaktnih podatkov, ampak le izkušnje s podobnimi pojavi,
- Kvantitativne metode so formalni postopki, ki predpostavljajo vztrajnost poslovnih procesov in uporabljajo podatke o pojavu iz preteklosti ter matematične modele za napoved;
  - ekstrapolacijske metode za napovedovanje endogenih pojavov (npr. različne metode povprečij),
  - korelacijske metode za napovedovanje eksogenih pojavov (npr. različne variante regresijske analize).

Časovne vrste

- zlasti nestacionarni pojavi se spreminjajo s časom:
  - podatke o njih statistično zajemamo v odvisnosti od časa,
  - v enakih diskretnih časovnih razmikih,

- take zapise statističnih podatkov o pojavu imenujemo časovne vrste ('time series') podatkov.

- časovna vrsta je tako niz istovrstnih podatkov (npr. o prodaji izdelka x), ki se nanašajo na zaporedne (sukcesivne) časovne razmike in nam posreduje sliko dinamike pojava; poznamo:
  - momentne časovne vrste predstavljajo presek stanja nekega pojava v določenem trenutku (npr. stanje zaloge artikla na prvi dan v mesecu),
  - intervalne časovne vrste predstavljajo gibanje pojava v določenih časovnih obdobjih (npr. prodaja nekega artikla v posameznih mesecih).

Časovne vrste

- V časovnih vrstah je čas neodvisna spremenljivka (x) :
  - čas lahko samo napreduje, vračanje v preteklost ni mogoče, časovne vrste so torej progresivne;
  - smatramo, da se čas kot neodvisna spremenljivka spreminja vedno v enakih diskretnih intervalih: leta, meseci, tedni, dnevi, ure ...

- Odvisna spremenljivka (y) je vrednost podatka o opazovanem pojavu,
  - vsaki vrednosti časa – neodvisne spremenljivke (x) je dodana ena in samo ena vrednost opazovanega pojava – odvisne spremenljivke (y)
  - med časom in obravnavanim pojavom torej obstoji neka **povezava** - soodvisnost, korelacija. Povezava pa ima lahko različne oblike.

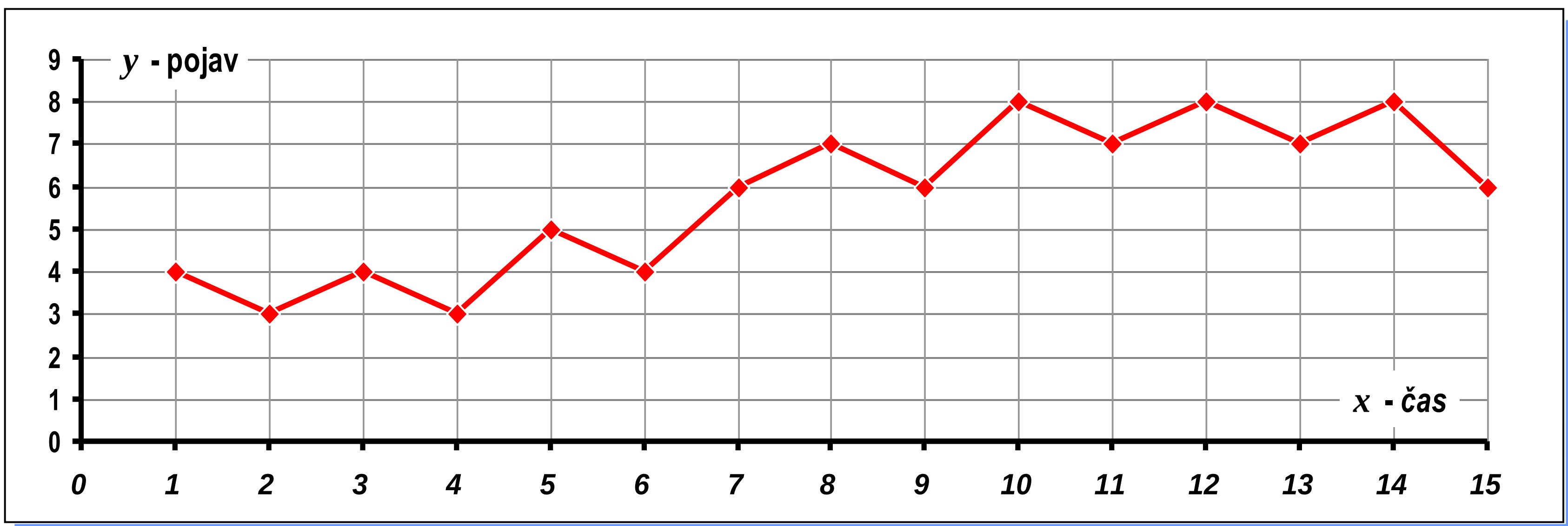
Časovne vrste

- Časovne vrste običajno <u>zapišemo v tabelarični obliki</u>, za lažje razumevanje in interpretacijo pa jih predstavimo lahko <u>tudi v obliki grafa</u>:

npr. leto 12 mesecev

naslednje leto teče od 13 dalje!

x - čas	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	...
y - pojav	4	3	4	3	5	4	6	7	6	8	7	8	7	8	6	...

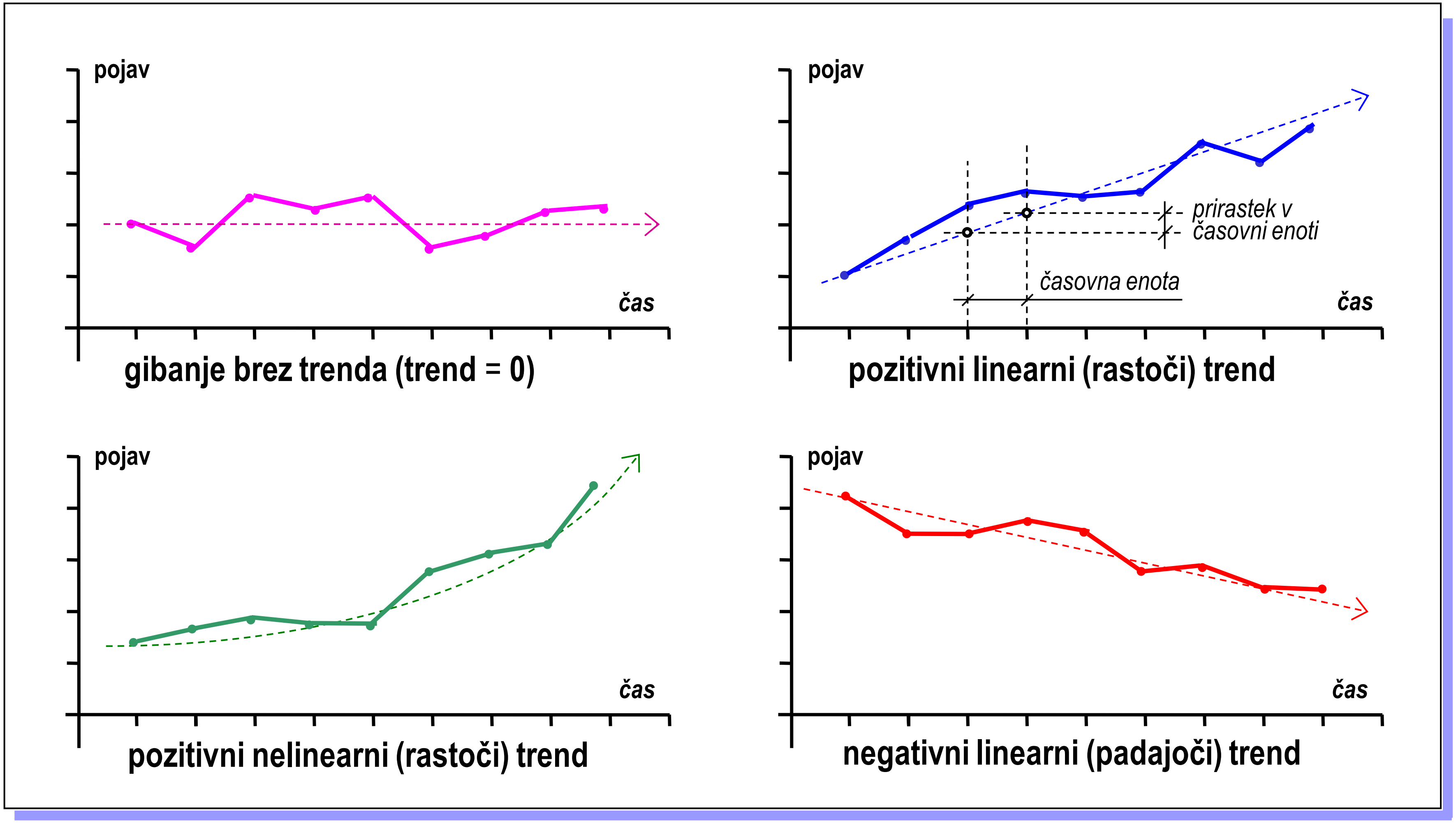


Povezava med časom in pojavom

- Funkcijska povezava obstoji, kadar je vrednost odvisne spremenljivke y (pojav) le funkcija časa - neodvisne spremenljivke x: $y = f(x)$, kar pa nastopa le redko;
- običajno je med časom in vrednostjo pojava stohastična povezava : $y = f(x) + \varepsilon$, kjer na vrednost odvisne spremenljivke y poleg neodvisne spremenljivke x vplivajo še slučajni, individualni vplivi ε, za katere pa veljajo zakoni verjetnosti. Ti vplivi so:
  - trend (T),
  - ciklični vplivi (C),
  - sezonski vplivi (S) in/ali periodični vplivi (P),
  - iregularitete - slučajnosti (I).

Trend

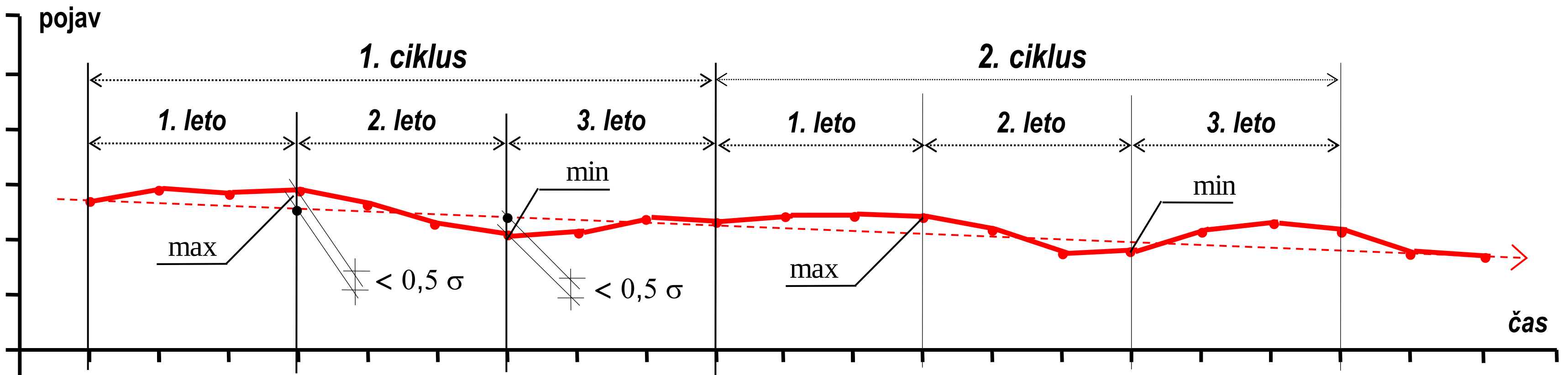
- podaja osnovno smer gibanja pojava, tolmači se kot prirastek/upadek (npr. pri linearni regresiji: b) osnovne vrednosti pojava (a) v časovni enoti (x):



Ciklična gibanja

- Vrednosti odvisne spremenljivke v daljšem obdobju (3 do 7 let) nihajo okrog osnovnega trenda:
  - ciklus nihanja je več kot eno leto,
  - maksimum in minimum sta vedno v istem obdobju ciklusa,
  - odstopanje minimuma / maksimuma od povprečja je $\leq \pm 0{,}5\,\sigma$.

- Ciklična gibanja so pogojena predvsem z dolgoročnimi gospodarskimi gibanji, pogosto pa tudi z modo.

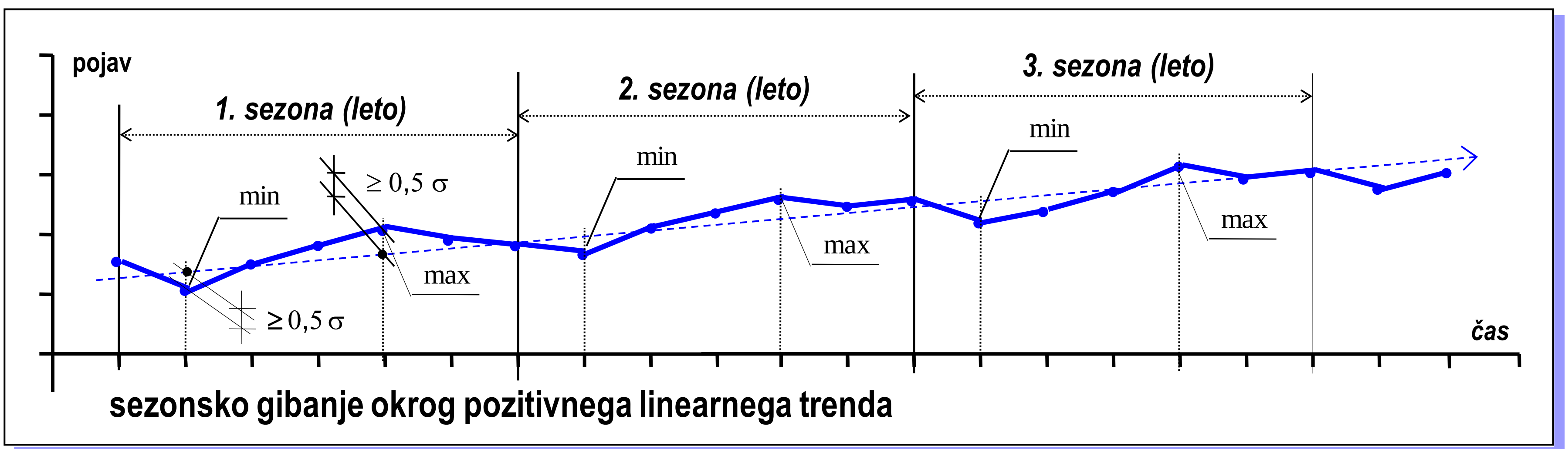


ciklično gibanje s ciklusom 3 leta okrog negativnega linearnega trenda

Sezonska in periodična gibanja

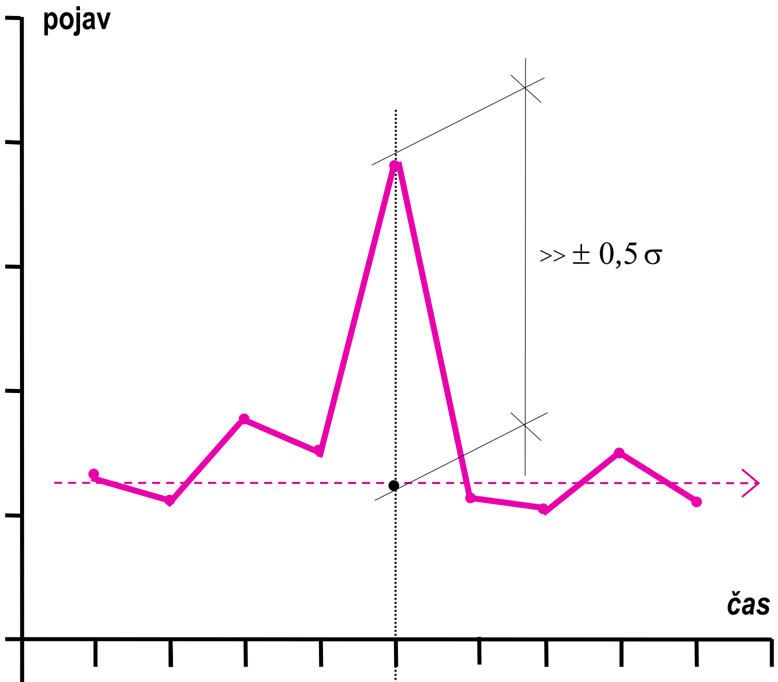
- Podobna so cikličnim, a se njihova dinamika kaže v krajših časovnih obdobjih:
  - ciklus nihanja je za sezonske vplive eno leto,
    za periodične vplive največkrat en mesec
  - maksimum in minimum sta vedno v istem obdobju ciklusa
  - odstopanje minimuma oziroma maksimuma od povprečja je $\geq \pm 0{,}5\ \sigma$

- sezonski vplivi so posledica klimatskih razmer ali mode, periodični pa drugih, redno ponavljajočih se dogodkov.



Iregularni - naključni vplivi

- Pojavljajo se sporadično, naključno;
  - njihov nastop je težko ali nemogoče predvideti,
  - njihovo odstopanje od povprečja (trenda) je $>> \pm 0{,}5\ \sigma$.



gibanje brez trenda z iregulariteto

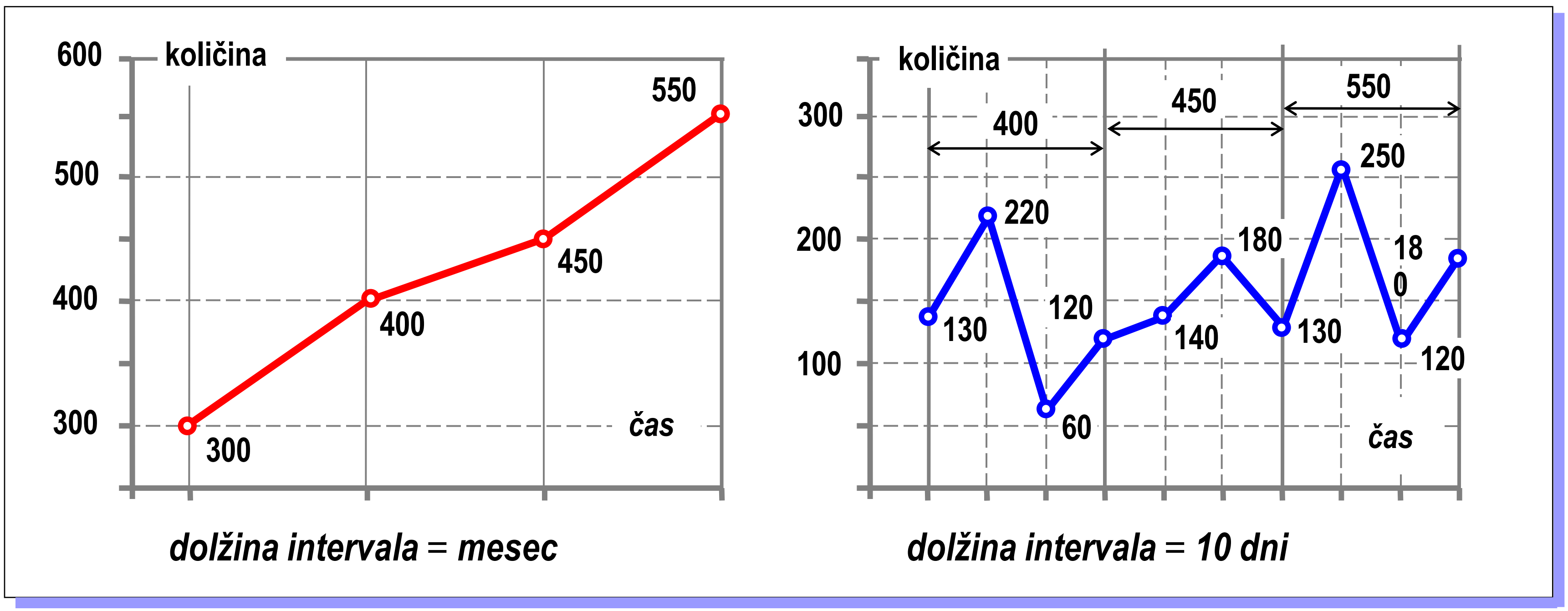
- epizodični (E) so vplivi, katerih pojav in učinek lahko obrazložimo,
  - takega značaja so slučajni, enkratni
    in kratkotrajni pojavi;
- naključni (N) so vplivi, katerih vzrokov ni mogoče ugotoviti,
  - nikakor jih ni mogoče predvideti
    in načeloma trajajo dalj časa.

Količina podatkov v časovnih vrstah

- o pojavu, ki ga opazujemo in želimo napovedovati:
  - za manj pomembne napovedi je potrebnih vsaj 7 podatkov iz preteklosti, (običajno se zahteva 12 podatkov)
  - za pojave sezonskega značaja morajo biti zagotovljeni podatki za vsaj dve sezoni (največkrat dve leti),
  - če se želi iz pojava izključiti iregularitete, pa se potrebuje podatke za najmanj štiri sezone.
- Če je podatkov preveč, uporabljamo drseče časovne vrste z omejenim številom členov:
  - časovna vrsta ima vseskozi enako število členov; število členov predstavlja tudi interval drsenja;
  - ko pride nov statistični podatek, se vrsto pomakne v levo: prvi, najstarejši podatek (na začetku vrste) izpade in ga nadomesti dotedanji drugi podatek, drugega tretji itd., na izpraznjeno mesto zadnjega, časovno najmlajšega podatka v vrsti pa se zabeleži novodošli podatek;
  - POZOR: to lahko zabriše pregled nad dolgoročnim gibanjem !

Časovni interval zajemanja podatkov

- vedno enakih časovnih intervalih - običajno teden, dekada ali mesec,
  - za dinamične pojave so ti intervali krajši, za umirjene pojave so lahko daljši,
  - predolga obdobja skrijejo iregularitete, pri prekratkih obdobjih iregularitete zabrišejo osnovno gibanje.



Kvaliteta podatkov v časovnih vrstah

- Vsi podatki o opazovanem pojavu
  - morajo biti vsebinsko enaki in primerljivi,
  - zajeti vedno na enak način in po enaki metodi.
- vedno se uporablja količinske podatke in iz njih izvede vrednostne (npr. napoved finančnih prihodkov iz prodaje).
- ne sme manjkati več kot 5% statističnih podatkov;
- ne smejo manjkati podatki za več zaporednih obdobij;
- če posamezni podatki izrazito odstopajo od ostalih, je treba preveriti, ali gre za iregularitete ali pa je prišlo do napake pri zajemanju podatkov;
- dvomljivih podatkov (ki se jih ne da pojasniti) se ne upošteva pri izračunu napovedi - nadomestijo se kot manjkajoč podatek:
  - Če podatek za neko obdobje manjka, se ga zasilno nadomesti ali z aritmetično sredino sosednjih dveh podatkov ali s srednjo vrednostjo vseh podatkov.
  - Če se dvomi v kakovost razpoložljivih podatkov, se jih neuporabi za napovedovanje; takrat se raje poslužujemo heurističnega ocenjevanja.

Urejanje podatkov v časovnih vrstah

- Če želimo podatke iz časovne vrste uporabili za ugotavljanje zakonitosti gibanja in napoved pojava v prihodnosti, je treba časovno vrsto najprej urediti:
  - analizirati, ugotoviti in pojasniti individualne vplive na gibanje opazovanega pojava;
- to se izvede lahko:
  - enostavno z oceno grafa časovne vrste,
  - ali **analitično**.

- Individualni vplivi zameglijo osnovno sliko gibanja pojava, kar povzroča napake pri napovedovanju, zato jih je treba pred ugotavljanjem zakonitosti gibanja pojava izločiti;
  - odstranjujejo se ciklični, sezonski in periodični vplivi ter iregularitete;
- trenda se ne odstranjuje, saj prav trend podaja osnovno značilnost gibanja pojava.

PRIMER UREJANJA IN ANALIZE ČASOVNE VRSTE:

Podatki o prodaji nekega izdelka po mesecih v zadnjih dveh letih :

mesec	1	2	3	4	5	6	7	8	9	10	11	12
prodaja *leto-1*	64	65	72	87	97	103	**108**	113	110	104	88	80
prodaja *leto*	82	83	90	108	48 **118**	128	134	137	132	123	110	98

Želimo ugotoviti značaj pojava in morebitne individualne vplive na njegovo gibanje?

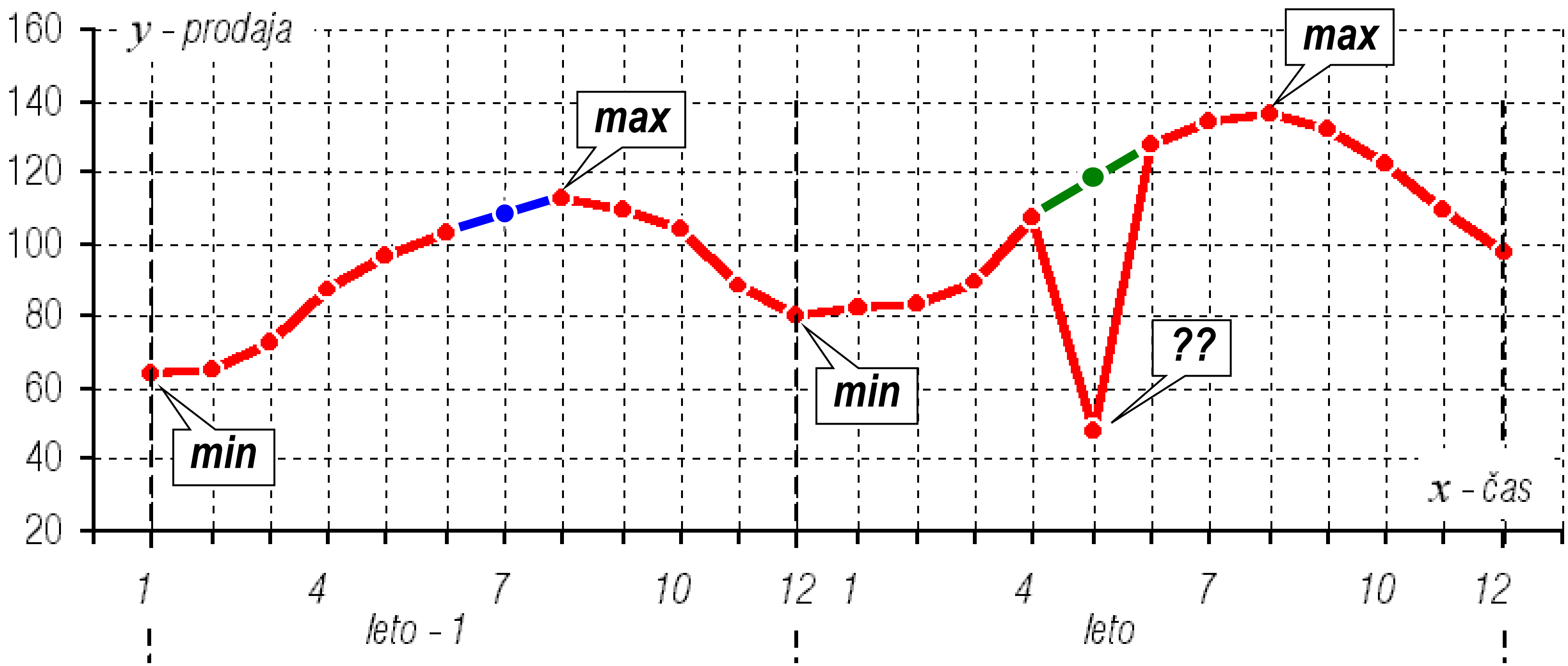
1. Podatek za julij predzadnjega leta manjka; nadomestimo ga s srednjo vrednostjo sosednjih dveh podatkov: (103+113)/2 =**108**.

2. Izračunamo povprečje časovne vrste in standardni odklon:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{2364}{24} = 98{,}5 \qquad \sigma_e = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}} = 23{,}6$$

3. Podatek za maj zadnjega leta izrazito odstopa od povprečja: **48** - 98,5 = –50,5 oz. /23,6 = **–2,14σ** (>> ± 0,5 σ). *Smemo sklepati, da je v tem mesecu prišlo do nekega iregularnega vpliva na prodajo ali je podatek napačen, zato ga lahko nadomestimo s srednjo vrednostjo sosednjih podatkov:* (108+128)/2 =**118** .

Za grobo oceno oblikujemo graf gibanja prodaje.



Manjkajoči <u>podatek za julij</u> predpreteklega leta nadomestimo z aritmetično sredino sosednjih dveh podatkov.

<u>Podatka za maj</u> preteklega leta v kasnejših računih ne upoštevamo, pač pa ga prav tako nadomestimo z aritmetično sredino sosednjih dveh podatkov.

Že površen pogled kaže, da bi <u>gibanje lahko imelo sezonski značaj</u> z maksimumom <u>vsako leto v mesecu avgustu in minimumom v januarju</u>.

Urejanje podatkov v časovnih vrstah

- Za odstranjevanje sezonskih vplivov iz podatkov o pojavu (dekompozicijo - desezonalizacijo) se izračunavajo **sezonski indeksi**,
  - za vsak mesec v letu povedo, koliko dejanska vrednost pojava odstopa (v plus ali minus) od idealnega pojava brez sezonskih, periodičnih in iregularnih vplivov.
  - ko dejanske vrednosti pojava korigiramo (deljenje) s sezonskimi indeksi, se dobijo idealizirane (desezonalizirane) vrednosti pojava, s katerimi se oblikuje model gibanja;
  - ko se na osnovi modela izvede napoved gibanja pojava v prihodnosti, se napovedane idealizirane vrednosti s sezonskimi indeksi (množenjem) vrnejo v realno napoved.

- Metode odstranjevanja sezonskih vplivov:
  - metoda verižnih indeksov,
  - metoda 12×12 centriranih povprečij.

Metoda 12×12 centriranih povprečij

① Potrebujemo podatke o pojavu (prodaji, povpraševanju) po mesecih za tri sezone (leta).

② Za vsak mesec v drugem letu izračunamo srednje vrednosti, povprečje vrednosti pojava za obdobje šest mesecev v nazaj in pet mesecev v naprej, torej za obdobje 12 mesecev:

$$\boxed{\bar{R}_n = \frac{\sum_{n-6}^{n+5} R_n}{12}}$$

R_n = vrednost pojava v nekem mesecu;
$\bar{R}_n$ = srednja vrednost (povprečje) vrednosti pojava za obdobje šest mesecev v nazaj in pet mesecev v naprej;
n = zaporedna številka meseca.

③ Izračunamo srednjo vrednost za dva sosednja meseca:

$$\boxed{\bar{R}'_n = \frac{\bar{R}_n + \bar{R}_{n+1}}{2}}$$

$\bar{R}'_n$ = srednja vrednost za dva sosednja meseca.

Metoda 12×12 centriranih povprečij

④ Vrednost podatka delimo z ustrezno srednjo vrednostjo za sosednja dva meseca, dobimo sezonski indeks za zadevni mesec:

$$\boxed{I_n = \frac{R_n}{\bar{R}'_n}}$$

I_n = *sezonski indeks za zadevni mesec.*

⑤ Vsota sezonskih indeksov za eno sezono (12 mesecev) mora biti <u>blizu 12</u>; odstopanja so lahko le zaradi zaokroževanja.

⑥ Desezonalizirane vrednosti pojava dobimo, če dejanske vrednosti delimo s sezonskim indeksom za zadevne mesece:

$$\boxed{R'_n = \frac{R_n}{I_n}}$$

R'_n = *desezonalizirana vrednost pojava v nekem mesecu.*

PRIMER DEKOMPOZICIJE ČASOVNE VRSTE Z METODO 12×12 CENTRIRANIH POVPREČIJ

① *V obliki časovne vrste imamo zabeležene podatke o prodaji nekega izdelka po mesecih v zadnjih treh letih:*

prodaja	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
leto-3	734	812	853	962	1078	933	827	902	411	392	332	685
leto-2	775	892	815	1043	1156	982	817	792	623	485	501	1054
leto-1	983	1121	1234	1198	1102	915	883	804	697	594	552	939

Ker ima gibanje prodaje evidentno sezonski značaj, želimo sezonske vplive izločiti – izvesti desezonalizacijo.

② *Za določanje sezonske komponente izberemo podatke za leto-2. Najprej izračunamo 12×12 centrirana povprečja po mesecih za obdobje mesec – 6 do mesec + 5 :*

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
leto-3												
leto-2	767,7	766,8	757,7	775,3	783,1	797,2	827,9	845,3	864,3	899,2	912,2	907,7
leto-1	902,1											

③ Srednje vrednosti povprečij za mesec in mesec +1 (iz tabele ②):

	1	2	3	4	5	6	7	8	9	10	11	12
leto-3												
leto-2	767,3	762,3	766,5	779,2	790,1	812,5	836,6	854,8	881,8	905,7	909,9	904,9
leto-1												

$$\bar{R}_2' = \frac{\bar{R}_2 + \bar{R}_3}{2} = \frac{766{,}8 + 757{,}7}{2} = \frac{1524{,}5}{2} = 762{,}3$$

$$\bar{R}_1' = \frac{\bar{R}_1 + \bar{R}_2}{2} = \frac{767{,}7 + 766{,}8}{2} = \frac{1534{,}5}{2} = 767{,}3$$

④ Sezonski indeksi (l-2 ① / ③):

Σ	1	2	3	4	5	6	7	8	9	10	11	12
12,09	1,01	1,17	1,06	1,34	1,46	1,21	0,9[illegible]	[illegible]	[illegible]	[illegible]	[illegible],55	1,16

$$I_2 = \frac{R_2}{\bar{R}_2'} = \frac{892}{762{,}3} = 1{,}17$$

$$I_1 = \frac{R_1}{\bar{R}_1'} = \frac{775}{767{,}3} = 1{,}01$$

⑥ S sezonskimi indeksi korigirane [illegible]lišče vrednosti:

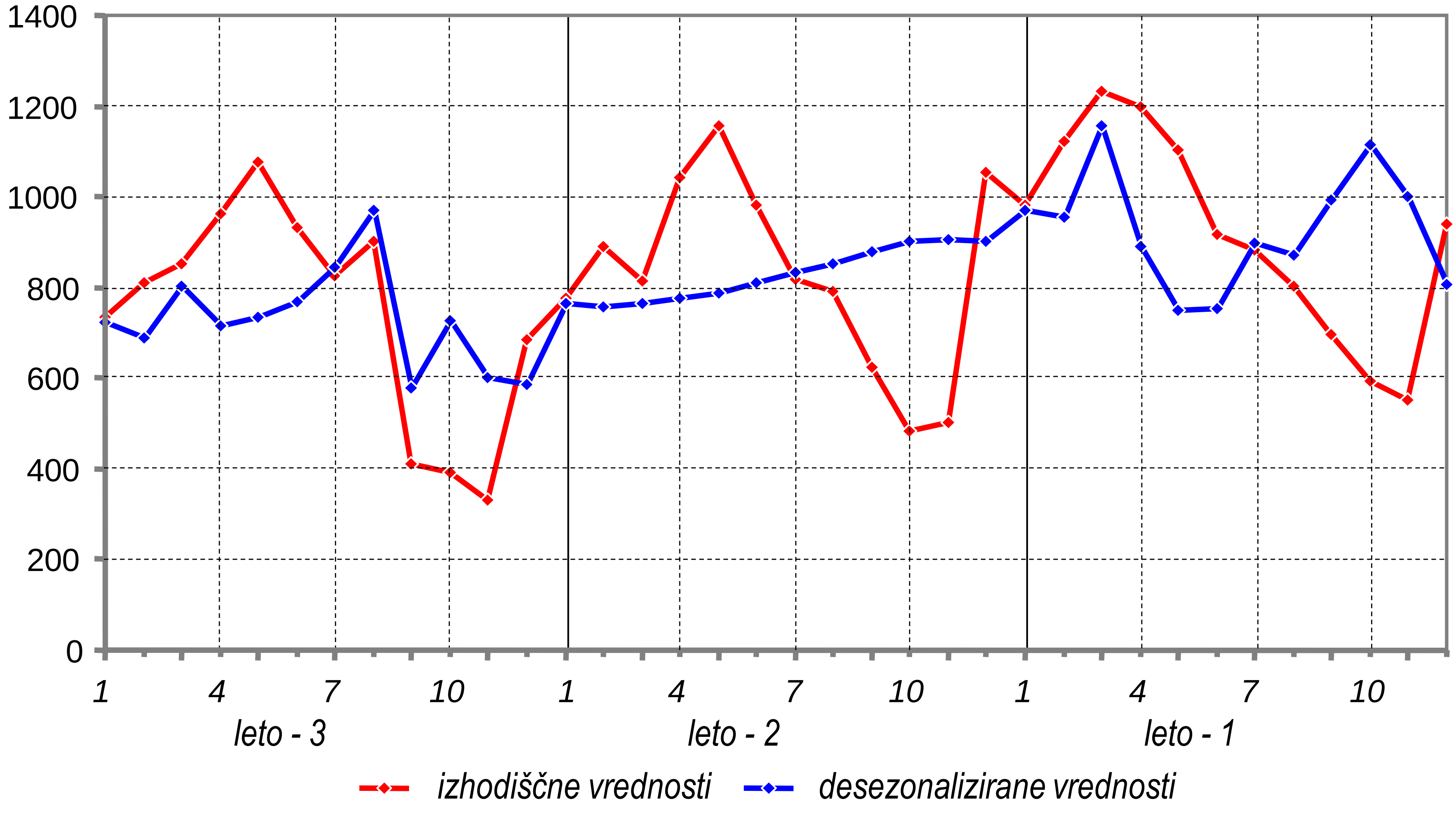
	1	2	3	4	5	6	7	8	9	10	11	12
leto-3	727	694	805	718	738	771	847	974	582	732	603	588
leto-2	767	762	767	779	790	813	837	855	882	906	910	905
leto-1	973	9[illegible]	1161	895	7[illegible]	757	901	874	996	1120	1004	809

= 812 /1,17 = 694

= 734 /1,01 = 727

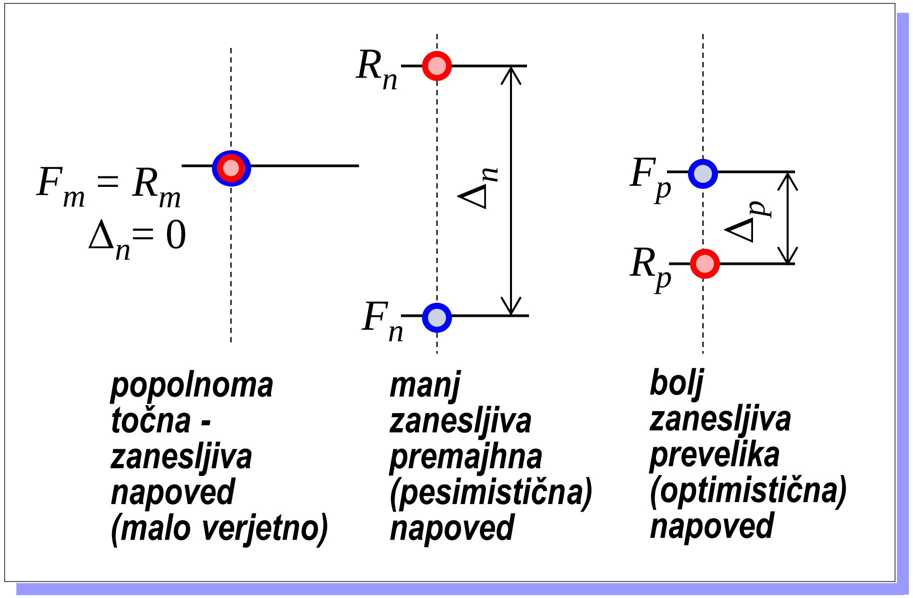
= 983 /1,01 = 973

Za vizualno kontrolo prikažemo izhodiščne vrednosti in desezonalizirane vrednosti podatkov v obliki grafa:



Zanesljivost napovedovanja

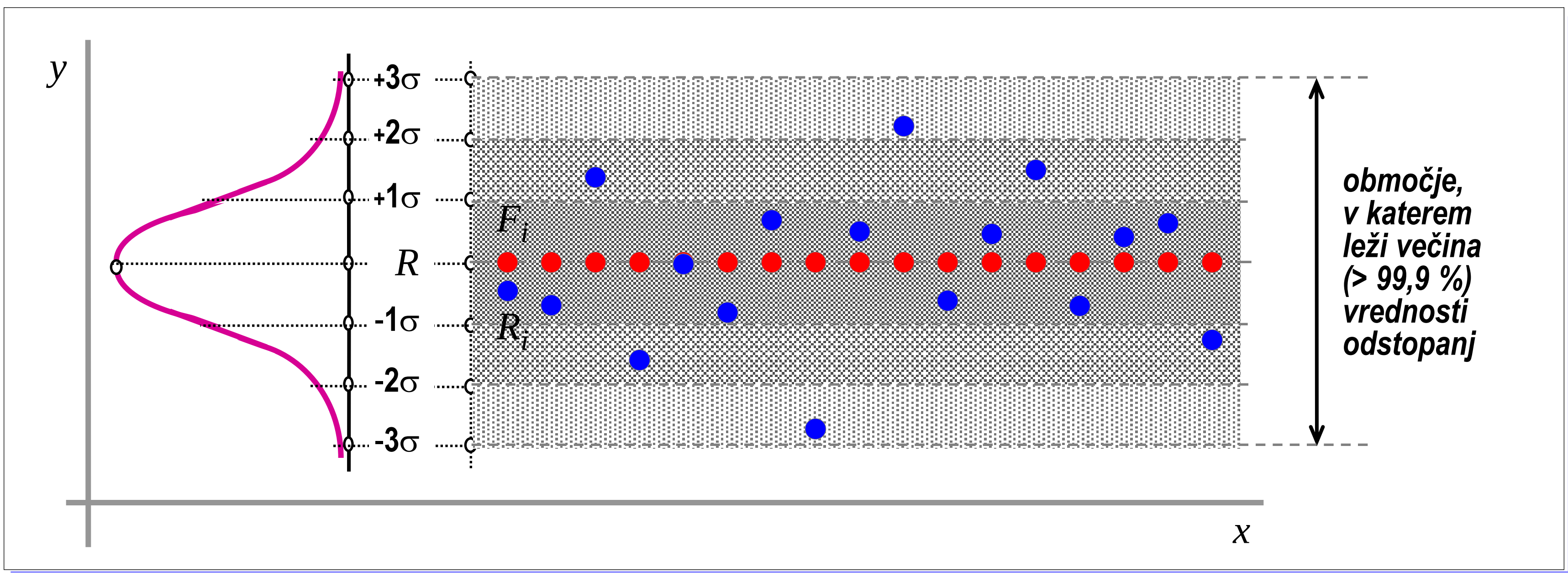
- Napoved naj bi čimbolj točno predvidela dejanski dosežek, naj ne bi bila niti previsoka, optimistična – niti prenizka, pesimistična;
  - vendar pa je malo verjetno, da bi napoved bila točno enaka dejanskemu dosežku.
- Merilo točnosti napovedovanja:
  - pove, koliko se dejansko dosežene vrednosti (R) razlikujejo od napovedi (F)



  - oz. kakšna je napaka (Δ_n) - odstopanje dejansko doseženih vrednosti od napovedi.

Statistična zanesljivost napovedovanja

- Mnogo napak - odstopanj (v plus ali v minus) je majhnih - napovedi so blizu dejanskim dosežkom, malo odstopanj pa je velikih:
  - Lahko predpostavimo, da je velikost odstopanj porazdeljena po zakonitostih normalne (Gaußove) porazdelitve.



Merila točnosti napovedovanja

- povprečni odklon napovedi oziroma srednja (povprečna) napaka napovedi
  (ME = 'Mean Error'),
- absolutni povprečni odklon napovedi oziroma absolutna srednja (povprečna) napaka napovedi
  (MAD = 'Mean Absolute Deviation'),
- odstotni (procentualni) absolutni povprečni odklon napovedi oziroma absolutna odstotna srednja (povprečna) napaka napovedi
  ($MAPE$ = 'Mean Absolute Percent Error')
- sledilni signal ('Tracking Signal').

Zanesljivost - točnost napovedovanja se ugotavlja vedno retrogradno (za nazaj) in se smatra, da velja za nekaj časa v naprej.

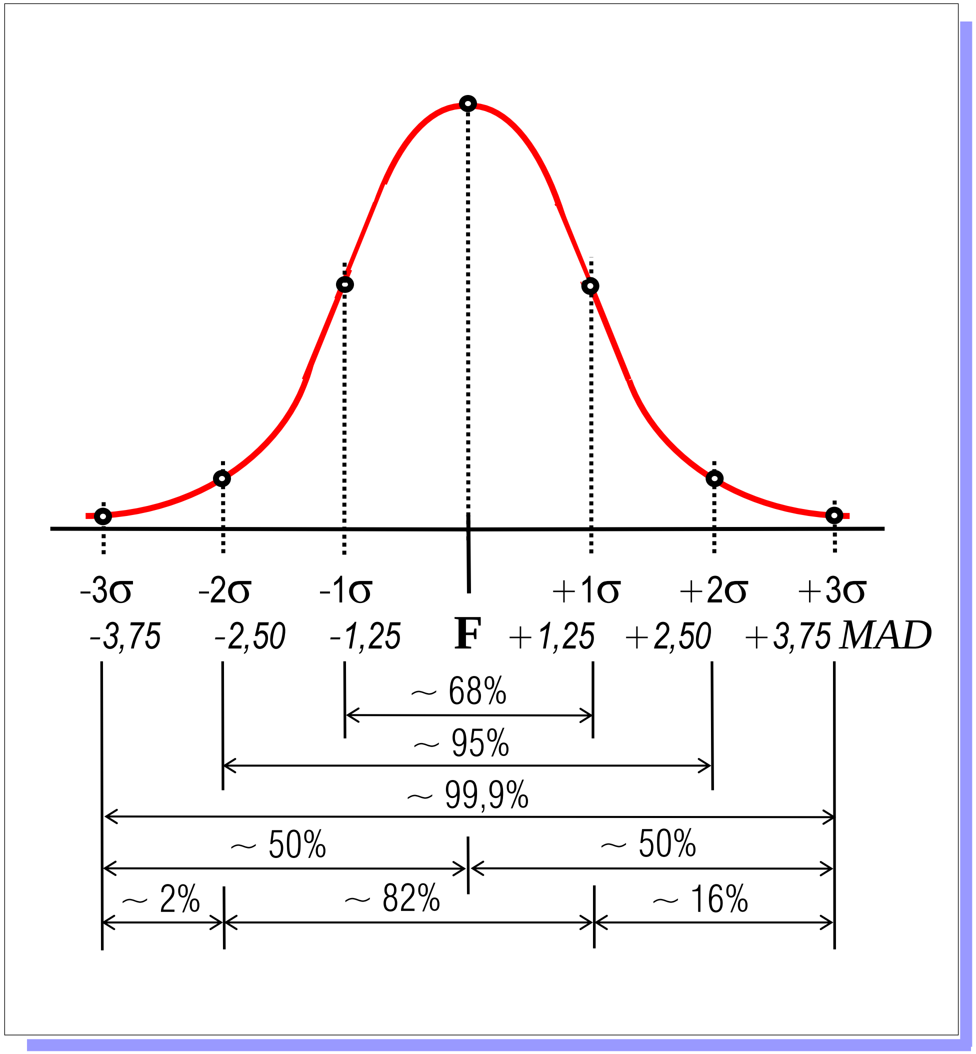
Absolutna srednja napaka napovedi (*MAD*)

$$MAD = \frac{\sum_{i=1}^{n} \left| (R_i - F_i) \right|}{n}$$

MAD = *absolutna srednja napaka napovedi*
R = *dejanska vrednost dogodka,*
F = *napovedana vrednost dogodka,*
i = *indeks (preštevno število) dogodkov,* $i = 1 .. n$
n = *število opazovanih dogodkov.*

- *MAD* predstavlja velikost povprečne napake napovedi ne glede na to, ali je le-ta pozitivna (premajhna) ali negativna (prevelika napoved).
- *MAD* se ne sme zamenjavati z varianco (σ^2) oziroma standardnim odklonom (σ):
  - če se operira z dovolj velikim vzorcem in se porazdelitev odstopanj podreja zakonitostim normalne porazdelitve, se sme privzeti, da je $\sigma = 1{,}25\ MAD$ oziroma $MAD = 0{,}8\ \sigma$,
  - in se lahko določi verjetnost, da bo napoved realizirana, (območje vrednosti)

Absolutna srednja napaka napovedi (*MAD*)



- ~ 68% verjetnosti je, da bo dejanska vrednost dogodka ležala med (točkovno) napovedano vrednostjo minus 1,25 *MAD* (oziroma –1 σ) in plus 1,25 *MAD* (oziroma +1 σ),
- ~ 95% verjetnosti je, da bo dejanska vrednost dogodka ležala med napovedano vrednostjo minus 2,5 *MAD* (–2 σ) in plus 2,5 *MAD* (+2 σ),
- > 99% verjetnosti je, da bo dejanska vrednost dogodka ležala med napovedano vrednostjo minus 3,75 *MAD* (–3 σ) in plus 3,75 *MAD* (+3 σ).

Pri napovedovanju vedno navajamo, s kakšno verjetnostjo leži napoved med zgornjo in spodnjo vrednostjo.

PRIMER DOLOČANJA ABSOLUTNE SREDNJE NAPAKE NAPOVEDI:

Desetkrat smo napovedali količino prodaje nekega izdelka in vsakič tudi ugotovili dejansko realizirano količino prodaje. Absolutna srednja napaka napovedi :

i	1	2	3	4	5	6	7	8	9	10	11
realizacija R_i	20	30	28	22	30	20	25	18	32	26	
napoved F_i	22	28	28	24	27	24	20	24	28	28	25
razlika $\|R_i - F_i\|$	2	2	0	2	3	4	5	6	4	2	

$$MAD = \frac{|-2| + 2 + 0 + |-2| + 3 + |-4| + 5 + |-6| + 4 + |-2|}{10} = \frac{30}{10} = 3$$

Torej lahko trdimo, da vrednost enajste napovedi

z verjetnostjo ~ 0,68 (68%) leži med (25 – 1,25 · 3) = 21,25 ≈ 21
in (25 + 1,25 · 3) = 28,75 ≈ 29 enotami,

z verjetnostjo ~ 0,95 (95%) leži med (25 – 2,50 · 3) = 17,50 ≈ 18
in (25 + 2,50 · 3) = 32,50 ≈ 33 enotami,

z verjetnostjo ~ 0,999 (99,9%) leži med (25 – 3,75 · 3) = 13,75 ≈ 14
in (25 + 3,75 · 3) = 36,25 ≈ 36 enotami.

Absolutna odstotna srednja napaka napovedi (*MAPE*)

$$MAPE = \frac{100 \cdot \sum_{i=1}^{n} \frac{\left|(R_i - F_i)\right|}{R_i}}{n}$$

$MAPE$ = *absolutna odstotna srednja napaka napovedi*
R = *dejanska vrednost dogodka,*
F = *napovedana vrednost dogodka,*
i = *indeks (preštevno število) dogodkov,* $i = 1 .. n$
n = *število opazovanih dogodkov.*

- $MAPE$ predstavlja absolutno odstotno velikost povprečne napake napovedi, ne glede na to, ali je le-ta pozitivna (premajhna) ali negativna (prevelika napoved).
- $MAPE$ je neodvisna od reda velikosti vrednosti dogodkov in napovedi, zato je primerna tudi za medsebojno primerjavo različnih dogodkov in napovedi.

PRIMER DOLOČANJA ABSOLUTNE ODSTOTNE SREDNJE NAPAKE NAPOVEDI:

Desetkrat smo napovedali količino prodaje nekega izdelka in vsakič tudi ugotovili dejansko realizirano količino prodaje. Absolutna odstotna srednja napaka napovedi za zadevni izdelek :

i	1	2	3	4	5	6	7	8	9	10	11
realizacija R_i	20	30	28	22	30	20	25	18	32	26	
napoved F_i	22	28	28	24	27	24	20	24	28	28	26
razlika $\|R_i - F_i\|$	2	2	0	2	3	4	5	6	4	2	
$\frac{\|R_i - F_i\|}{R_i}$	0,1	0,07	0	0,09	0,1	0,2	0,2	0,33	0,13	0,33	

$$MAPE = \frac{100 \cdot (0{,}1 + 0{,}07 + 0 + 0{,}09 + 0{,}1 + 0{,}2 + 0{,}2 + 0{,}33 + 0{,}13 + 0{,}33)}{10} =$$

$$= \frac{155}{10} = 15{,}5\ \%$$

V povprečju so napovedi odstopale za 15,5 % .

Sledilni signal

- Za spremljanje napovedi skozi daljši čas se uporablja sledilni signal ('tracking signal'), ki meri, kako zanesljivo se napoveduje dejanske vrednosti;
- izračuna se kot drseča vsota napak napovedi, deljeno z absolutno povprečno napako napovedi:

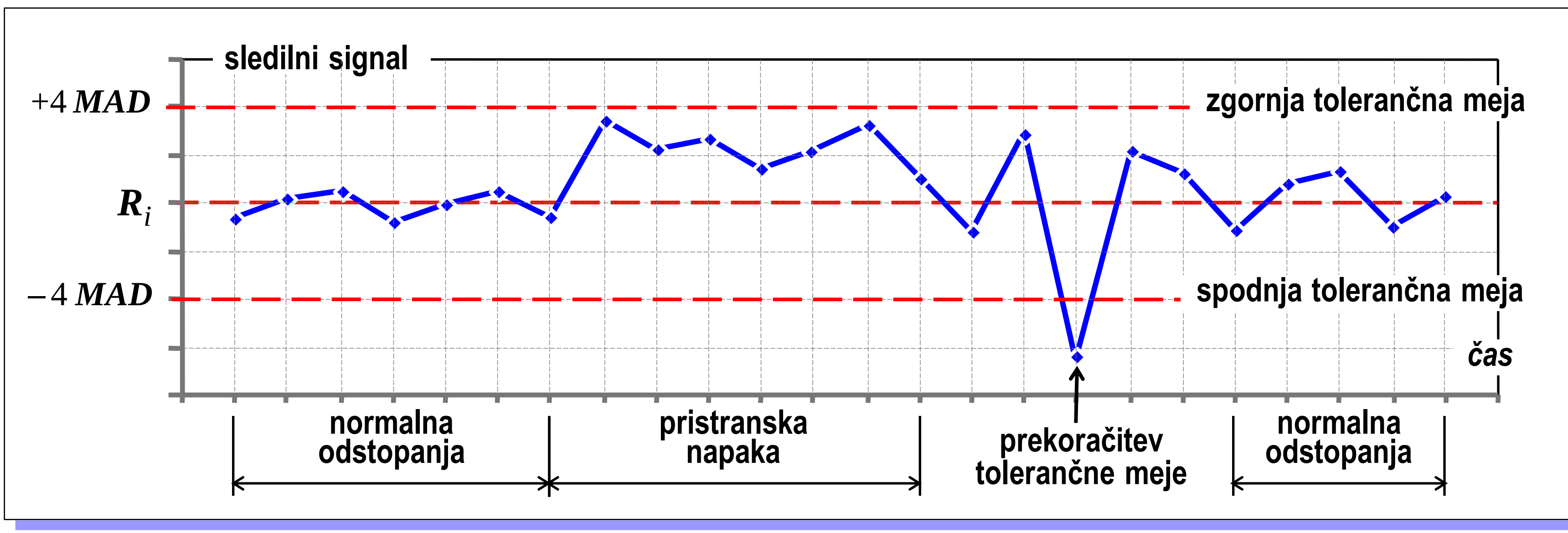
$$TS = \frac{\sum_{i=1}^{n}\left(R_i - F_i\right)}{MAD}$$

TS = *sledilni signal*
R = *dejanska vrednost dogodka,*
F = *napovedana vrednost dogodka,*
i = *indeks (preštevno število) dogodkov,* $i = 1 .. n$
MAD = *absolutna srednja napaka napovedi.*

- Pozitiven sledilni signal kaže na pesimistično napoved, negativen signal nasprotno pomeni, da je napoved optimistična;
- dobra napoved in s tem dober sledilni signal imata enako pozitivnih kot negativnih odstopanj.

Sledilni signal

- Sledilni signal se primerja z vnaprej določenimi tolerančnimi mejami;
- če prekorači zgornjo ali spodnjo mejo, to kaže na neprimerno metodo napovedovanja;
- stalna težnja, da napovedi zelo odstopajo od dejanskih vrednosti, pa se imenuje pristranska, subjektivna napaka ('bias').



PRIMER DOLOČANJA SLEDILNEGA SIGNALA:

Imamo podatke o napovedi povpraševanja in dejanskem povpraševanju za zadnjih šest obdobij.

Izračunati želimo sledilni signal in ugotoviti, ali se napovedi obnašajo primerno :

i	napoved F_i	realizacija R_i	$R_i - F_i$	$\Sigma\ R_i - F_i$	$\lvert R_i - F_i \rvert$	$\Sigma\ \lvert R_i - F_i \rvert$		MAD	sledilni signal TS	
1	100	90	−10	−10	10	10		10,0	−10/10 =	−1,0
2	100	95	−5	−15	5	15		7,5	−15/7,5 =	−2,0
3	100	115	+15	0	15	30	/ 3=	10,0	0/10 =	0
4	110	100	−10	−10	10	40		10,0	−10/10 =	−1,0
5	110	125	+15	+5	15	55		11,0	+5/11 =	+0,5
6	110	140	+30	+35	30	85		14,2	+35/14,2 =	+2,5

Ob koncu šestega obdobja je absolutna povprečna napaka napovedi (MAD) 14,2 in sledilni signal +2,5 MAD. Sledilni signal se je spreminjal od –2,0 MAD do +2,5 MAD.

Ekstrapolacijske metode

Enostavne srednje vrednosti

- Za napoved vrednosti nekega pojava v prihodnosti uporabimo srednjo vrednost (aritmetično sredino, povprečje) vseh podatkov o pojavu v preteklosti:

$$\boxed{F_{i+1} = \bar{R} = \frac{1}{n} \cdot \sum_{i=1}^{n} R_i}$$

F_{i+1} = napoved vrednosti pojava v naslednjem obdobju (F = 'forecast' – napoved pojava)

$\bar{R}$ = enostavna aritmetična sredina podatkov o pojavu

R_i = podatki o opazovanem pojavu v posameznih obdobjih, (R = 'reality' – dejanska vrednost pojava), $i = 1 .. n$

n = število obdobij, za katera imamo podatke.

  - Izračunava se v vsakem obdobju sproti, vedno se korigira z novimi podatki, <u>na napoved vplivajo vse vrednosti</u>, <u>tudi tiste iz preteklosti</u>, vpliv novejših podatkov je majhen.
  - Model gibanja pojava želi umiriti – <u>odzivnost, je slaba</u>, zato je uporabna le, kadar gre za <u>gibanja brez izrazitega trenda</u>, sezonskih vplivov ter iregularitet.

Ekstrapolacijske metode

Drseče srednje vrednosti

- V izračunu ne uporabljamo vseh statističnih podatkov, ki jih imamo na razpolago, pač pa le podatke za zadnjih nekaj obdobij (drseča časovna vrsta):

$$F_{i+1} = \bar{R}' = \frac{1}{m} \cdot \sum_{i=n-m+1}^{n} R_i$$

m = število členov drseče časovne vrste - število zadnjih obdobij, za katere bomo upoštevali podatke (interval drsenja)

vse ostalo isto, kot za enostavne srednje vrednosti

  - Drsečo aritmetično sredino se izračunava v vsakem obdobju sproti, vendar le s podatki za toliko zadnjih obdobij, kolikor je interval drsenja; zato na napoved vpliva le nekaj aktualnih vrednosti;
  - čim krajši je interval drsenja, toliko hitreje se model odziva na spremembe – vendar tudi na slučajne vplive.
  - model skuša gibanje pojava umiriti, a je bolj odziven, kot enostavne aritmetične sredine.

Ekstrapolacijske metode

Utežene srednje vrednosti

- Ideja uteževanja (ponderiranja):
  - drseča aritmetična sredina za npr. 4 obdobja:

$$F_5 = \frac{R_1 + R_2 + R_3 + R_4}{4} = \frac{1}{4} \cdot R_1 + \frac{1}{4} \cdot R_2 + \frac{1}{4} \cdot R_3 + \frac{1}{4} \cdot R_4 =$$
$$= 0,25 \cdot R_1 + 0,25 \cdot R_2 + 0,25 \cdot R_3 + 0,25 \cdot R_4$$

uteži posameznih podatkov

  - v tem primeru so vse uteži (pondri) enako velike,
  - vendar to ni nujno, uteži so lahko različno velike, a pod pogojem, da je njihova vsota enaka 1;
  - večja utež pomeni, da ima izbrani podatek večjo težo – bolj vpliva na napoved;
  - s tem pa dobimo orodje za vplivanje na napoved - odzivnost.

Ekstrapolacijske metode
Utežene srednje vrednosti

- Utežena (ponderirana) aritmetična sredina:

$$F_{i+1} = \bar{R}' = \frac{1}{m} \cdot \sum_{i=n-m+1}^{n} R_i = \sum_{i=n-m+1}^{n} \frac{1}{m} \cdot R_i = \sum_{i=n-m+1}^{n} q_j \cdot R_i$$

q_j = *ponder – utež posameznega podatka v časovni vrsti,* $j = 1 \,..\, m$

vse ostalo isto, kot za enostavne in drseče srednje vrednosti

- Vrednosti uteži posameznih podatkov ležijo med 0 in 1, v načelu so manjše od 1, vsota uteži mora biti enaka 1;
- če se da večjo težo starejšim podatkom (z začetka časovne vrste), se skuša napoved umiriti, če pa imajo večjo težo mlajši podatki (s konca časovne vrste), bo napoved bolj sledila novejšemu gibanju pojava.

PRIMER NAPOVEDI Z UTEŽENIMI SREDNJIMI VREDNOSTMI

Porabo materiala po prejšnjem primeru napovedujemo z uteženimi drsečimi aritmetičnimi sredinami z intervalom drsenja 5 obdobij, z dvema različnima vrstama uteži q'_j *in* q''_j *.*
Izračunane podatke prikažemo tabelarično in grafično.

Uteži (pondri):

j	1	2	3	4	5	Σq_j
q'_j	**0,30**	**0,25**	**0,20**	**0,15**	**0,10**	**1,00**
q''_j	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**1,00**

Vrsta uteži q'_j *daje večjo težo starejšim podatkom, medtem ko vrsta uteži* q''_j *poudarja novejše podatke.*

Napoved z vrsto uteži q'_j , ki daje večjo težo starejšim podatkom:

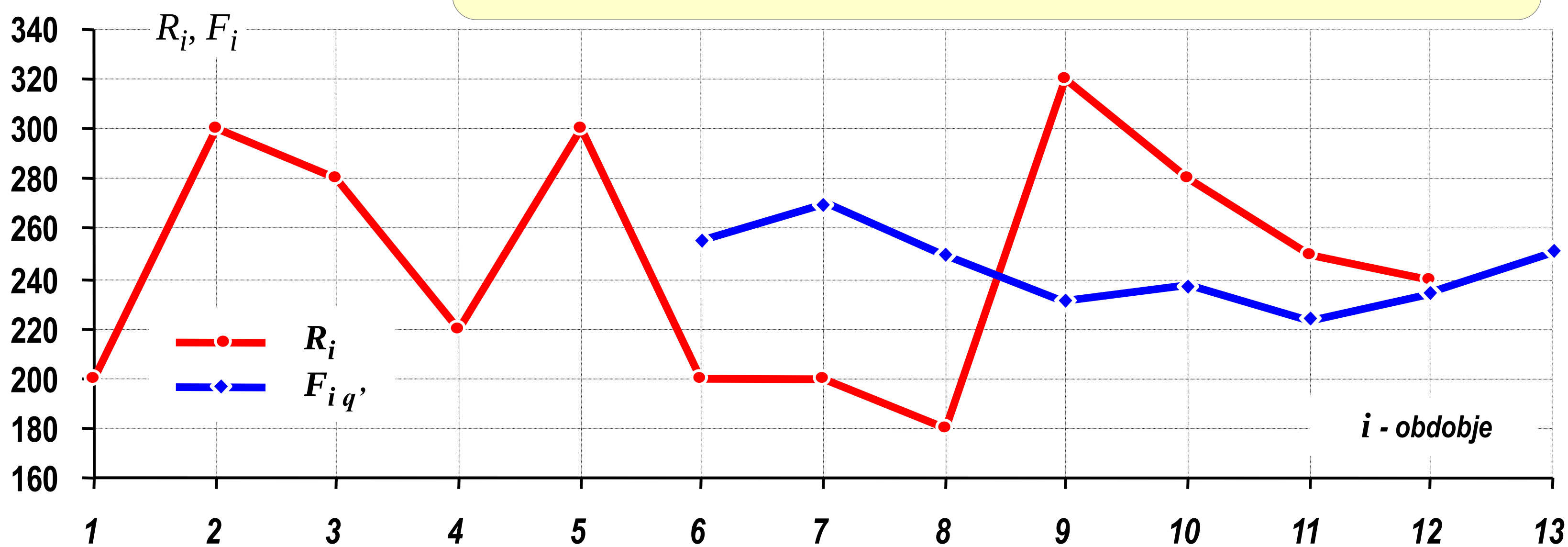
i	1	2	3	4	5	6	7	8	9	10	11	12	13
R_i	200	300	280	220	300	200	200	180	320	280	250	240	
F_i za q'_j						254	269	249	229	239	222	236	251

$$F_6 = q'_1 \cdot R_1 + q'_2 \cdot R_2 + q'_3 \cdot R_3 + q'_4 \cdot R_4 + \ldots =$$

$= 0{,}30 \cdot 200 + 0{,}25 \cdot 300 + 0{,}20 \cdot 280 + 0{,}1 \ldots 300 = 254$

$$F_7 = q'_1 \cdot R_2 + q'_2 \cdot R_3 + q'_3 \cdot R_4 + q'_4 \cdot R_5 + q'_5 \cdot R_6 =$$

$= 0{,}30 \cdot 300 + 0{,}25 \cdot 280 + 0{,}20 \cdot 220 + 0{,}15 \cdot 300 + 0{,}10 \cdot 200 = 269$



Absolutna srednja napaka napovedi (MAD):

i	1	2	3	4	5	6	7	8	9	10	11	12	13
R_i	200	300	280	220	300	200	200	180	320	280	250	240	
F_i						254	269	249	229	239	222	236	251
$\lvert R_i - F_i \rvert$						54	69	69	91	41	28	16	

Število statističnih podatkov = 7

$$MAD = \frac{54 + 69 + 69 + 91 + 41 + 28 + 16}{7} = \frac{368}{7} = 52{,}6$$

Lahko trdimo, da napoved porabe zadevnega materiala v trinajsti terminski enoti (tednu, mesecu ...)

z verjetnostjo ~ 0,68 (68%) leži med (251 – 1,25 · 52,6) ≈ 185
in (251 + 1,25 · 52,6) ≈ 317 enotami,

z verjetnostjo ~ 0,95 (95%) leži med (251 – 2,50 · 52,6) ≈ 120
in (251 + 2,50 · 52,6) ≈ 382 enotami,

z verjetnostjo ~ 0,999 (99,9%) leži med (251 – 3,75 · 52,6) ≈ 54
in (251 + 3,75 · 52,6) ≈ 448 enotami.

Napoved z vrsto uteži q''_j, ki daje večjo težo novejšim podatkom:

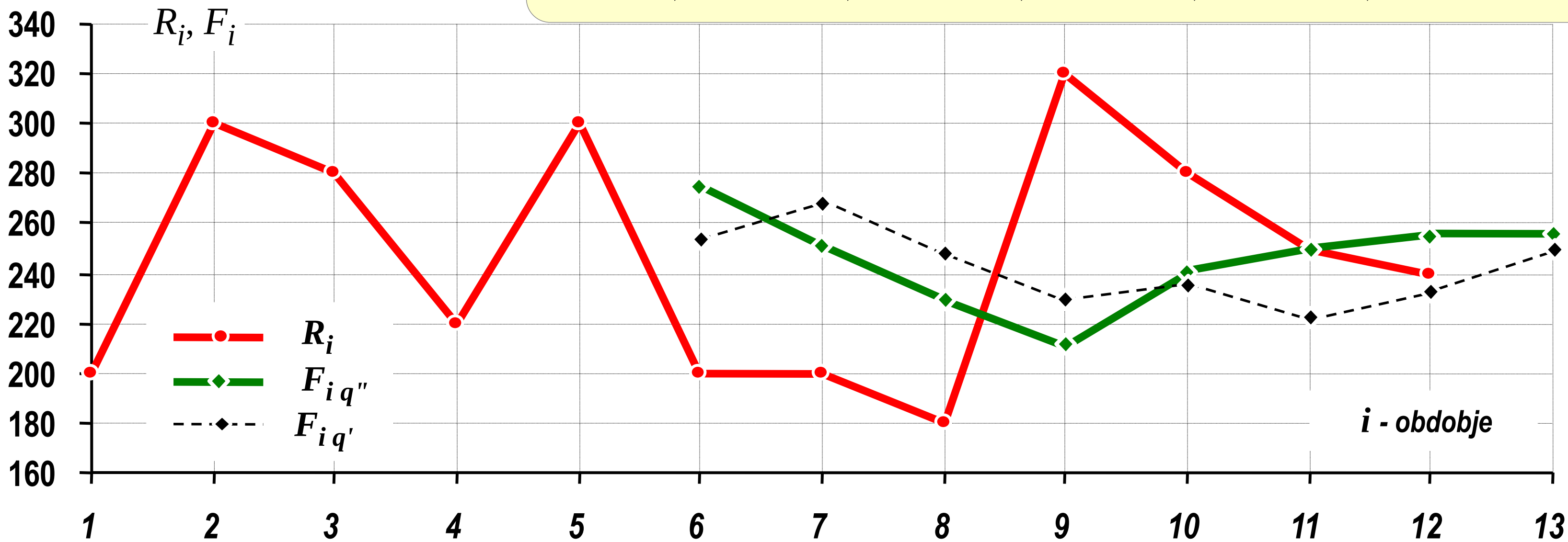
i	1	2	3	4	5	6	7	8	9	10	11	12	13
R_i	200	300	280	220	300	200	200	180	320	280	250	240	
F_i za q''_j						266	251	231	211	241	250	256	256

$F_6 = q''_1 \cdot R_1 + q''_2 \cdot R_2 + q''_3 \cdot R_3 + q''_4 \cdot R_4 + \ldots =$

$= 0{,}10 \cdot 200 + 0{,}15 \cdot 300 + 0{,}20 \cdot 280 + 0{,}25 \cdot 22\ldots = 266$

$F_7 = q''_1 \cdot R_2 + q''_2 \cdot R_3 + q''_3 \cdot R_4 + q''_4 \cdot R_5 + q''_5 \cdot R_6 =$

$= 0{,}10 \cdot 300 + 0{,}15 \cdot 280 + 0{,}20 \cdot 220 + 0{,}25 \cdot 300 + 0{,}30 \cdot 200 = 251$



Absolutna srednja napaka napovedi (MAD):

i	1	2	3	4	5	6	7	8	9	10	11	12	13
R_i	200	300	280	220	300	200	200	180	320	280	250	240	
F_i						256	251	231	211	241	250	256	256
$\lvert R_i - F_i \rvert$						44	51	51	109	39	0	16	

Število statističnih podatkov = 7

$$MAD = \frac{44 + 51 + 51 + 109 + 39 + 0 + 16}{7} = \frac{307}{7} = 44{,}3$$

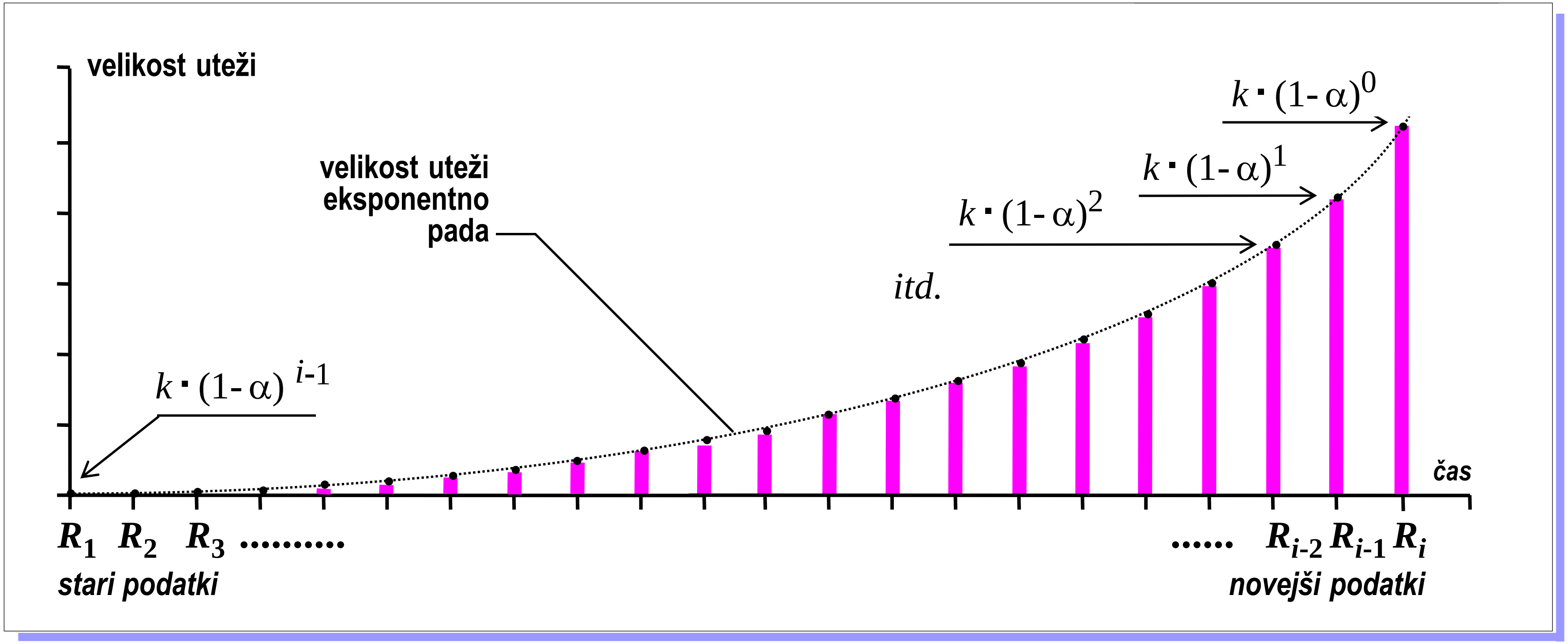
Lahko trdimo, da napoved porabe zadevnega materiala v trinajsti terminski enoti (tednu, mesecu ...)

z verjetnostjo ~ 0,68 (68%) leži med (256 – 1,25 · 44,3) ≈ 201
in (256 + 1,25 · 44,3) ≈ 311 enotami,
z verjetnostjo ~ 0,95 (95%) leži med (256 – 2,50 · 44,3) ≈ 145
in (256 + 2,50 · 44,3) ≈ 367 enotami,
z verjetnostjo ~ 0,999 (99,9%) leži med (256 – 3,75 · 44,3) ≈ 90
in (256 + 3,75 · 44,3) ≈ 422 enotami.

Ekstrapolacijske metode

Eksponentno glajenje ('Exponential Smothing')

- Ideja: avtomatizacija dodeljevanja uteži različno starim podatkom pri uteženi aritmetični sredini;
- če zmanjševanje uteži poteka eksponentno, govorimo o eksponentnem glajenju.



Ekstrapolacijske metode

Enostavno eksponentno glajenje 1. reda

- Formulacija:

$$F_{i+1} = F_i + \alpha \cdot (R_i - F_i)$$

kjer je:

F_{i+1} = *napoved vrednosti pojava v obdobju* $i+1$, $i = 1 .. n$

F_i = *predhodna napoved vrednosti pojava v obdobju* i, $i = 1 .. n$

R_i = *dejanska vrednost pojava v obdobju* i, $i = 1 .. n$

α = *konstanta (faktor) eksponentnega glajenja;* $\alpha = 0 .. 1$

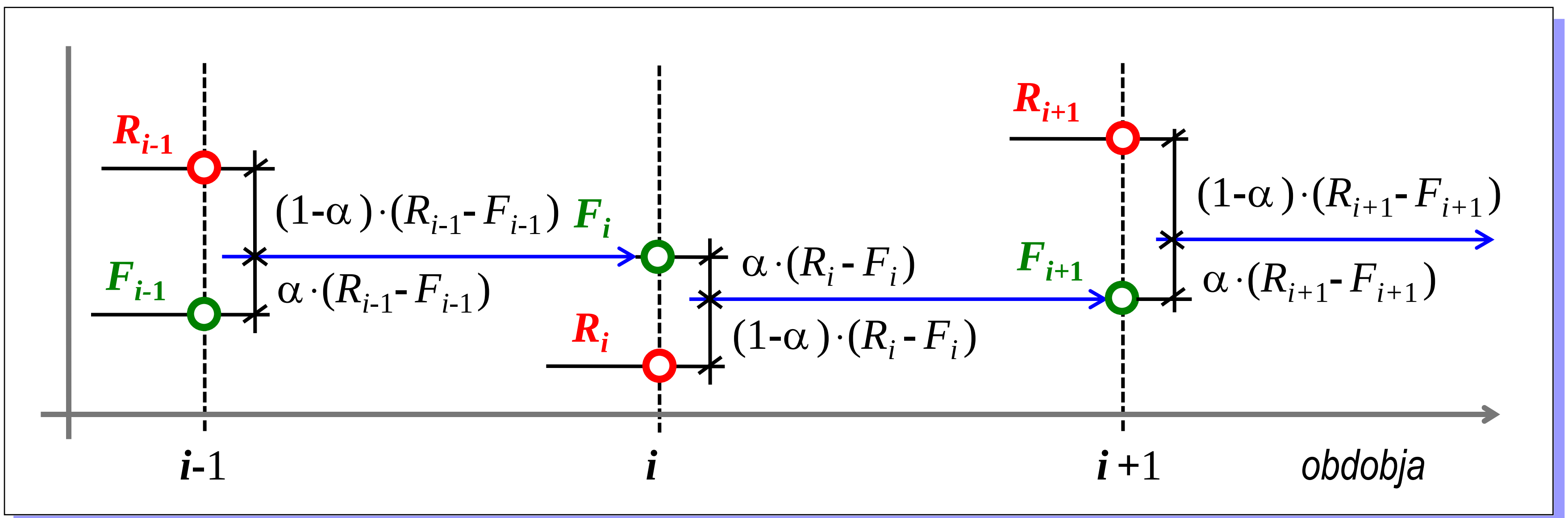
n = *število obdobij, za katera imamo podatke .*

- Velikost konstante α je v praksi med 0,1 in 0,25 do 0,33; z izbiro velikosti faktorja glajenja vplivamo na odzivnost napovedi .

Ekstrapolacijske metode

Enostavno eksponentno glajenje 1.reda

- Napoved za naslednje plansko obdobje
  - je enaka razliki med napovedjo za prejšnje obdobje in dejansko vrednostjo dogodka v tem obdobju, korigirani za določen del napake stare napovedi:



Ekstrapolacijske metode

Enostavno eksponentno glajenje 1. reda

- Čim manjši je α, tem večji vpliv na napoved imajo stari podatki, manjši α teži k umirjanju, stabilizaciji napovedi;
- večji α pomeni hitro prilagajanje napovedi najnovejšemu gibanju pojava (a tudi hitro odzivanje na iregularitete).
- Napoved z $\alpha = 0$ je prenos napovedi za predhodno obdobje v naslednje, napoved z $\alpha = 1$ pa prenese dejansko vrednost pojava v predhodnem obdobju v napoved v naslednjem obdobju.
- Izračun se lahko začne takoj, ne da bi bilo potrebno dolgo časa zbirati statistične podatke; za prvo napoved se enostavno predpostavi, da je $F_1 = R_1$.
- Enostavno eksponentno glajenje prvega reda je uporabno za napoved pojavov, ki nimajo trenda ali sezonskih vplivov.
- Slabost: zavajanje, da je vodenje statistike o pojavu nepotrebno.

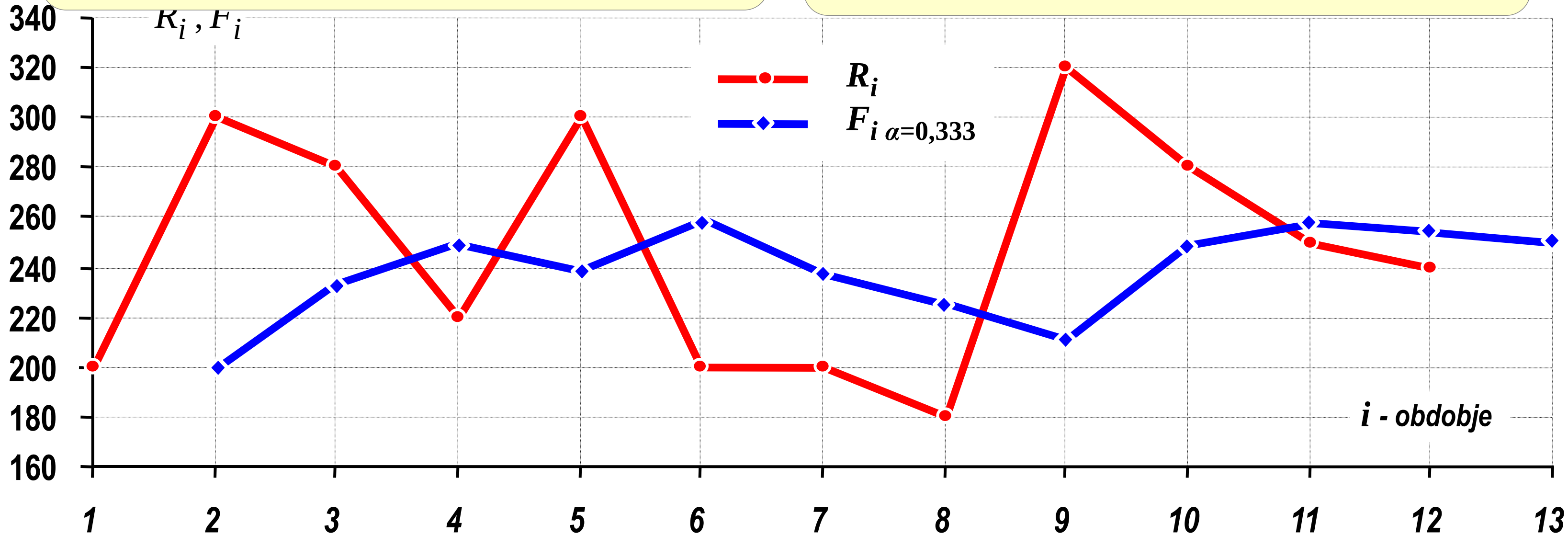
PRIMER NAPOVEDOVANJA Z METODO ENOSTAVNEGA EKSPONENTNEGA GLAJENJA PRVEGA REDA

Prodajo izdelka po prejšnjem primeru napovedujemo z enostavnim eksponentnim glajenjem prvega reda. Napoved z α = 0,333 (poudarek novejšim podatkom) :

i	1	2	3	4	5	6	7	8	9	10	11	12	13
R_i	200	300	280	220	300	200	200	180	320	280	250	240	
$F_{i,\alpha=0,333}$		200	233	249	239	259	239	226	211	247	258	255	250

$F_{2,\alpha=0,333} = F_1 + \alpha\cdot(R_1 - F_1) =$
$= 200+0,333\cdot(200-200) = 200$

$F_{3,\alpha=0,333} = F_2 + \alpha\cdot(R_2 - F_2) =$
$= 200+0,333\cdot(300-200) = 233$



Absolutna srednja napaka napovedi (MAD):

i	1	2	3	4	5	6	7	8	9	10	11	12	13
R_i	200	300	280	220	300	200	200	180	320	280	250	240	
F_i		200	233	249	239	259	239	226	211	247	258	255	250
$\lvert R_i - F_i \rvert$		100	47	29	61	59	39	46	109	33	8	15	

Število statističnih podatkov = 11

$$MAD = \frac{100 + 47 + 29 + 61 + 59 + 39 + 46 + 109 + 33 + 8 + 15}{11} = \frac{546}{11} = 49{,}6$$

Lahko trdimo, da napoved porabe zadevnega materiala v trinajsti terminski enoti (tednu, mesecu ...)

z verjetnostjo ~ 0,68 (68%) leži med (250 – 1,25 · 49,6) ≈ 188
in (250 + 1,25 · 49,6) ≈ 312 enotami,

z verjetnostjo ~ 0,95 (95%) leži med (250 – 2,50 · 49,6) ≈ 126
in (250 + 2,50 · 49,6) ≈ 374 enotami,

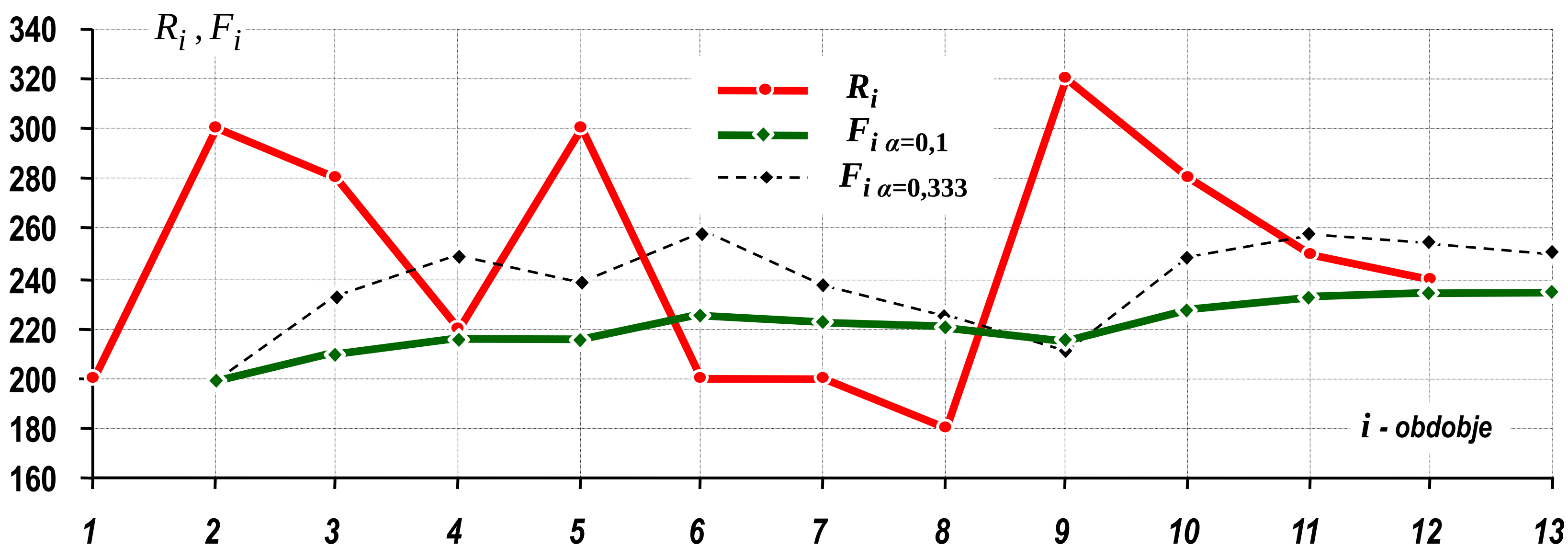
z verjetnostjo ~ 0,999 (99,9%) leži med (250 – 3,75 · 49,6) ≈ 64
in (250 + 3,75 · 49,6) ≈ 436 enotami.

Napoved z α = 0,1 (poudarek starejšim podatkom) :

i	1	2	3	4	5	6	7	8	9	10	11	12	13
R_i	200	300	280	220	300	200	200	180	320	280	250	240	
$F_{i,\alpha=0,1}$		200	210	217	217	225	223	221	217	227	232	234	235

$F_{2,\alpha=0,1} = F_1 + \alpha \cdot (R_1 - F_1) =$
$= 200+0,1 \cdot (200-200) = 200$

$F_{3,\alpha=0,1} = F_2 + \alpha \cdot (R_2 - F_2) =$
$= 200+0,1 \cdot (300-200) = 210$



Absolutna srednja napaka napovedi (MAD):

i	1	2	3	4	5	6	7	8	9	10	11	12	13
R_i	200	300	280	220	300	200	200	180	320	280	250	240	
F_i		200	210	217	217	225	223	221	217	227	232	234	235
$\lvert R_i - F_i \rvert$		100	70	3	83	25	23	41	103	53	18	6	

Število statističnih podatkov = 11

$$MAD = \frac{100 + 70 + 3 + 83 + 25 + 23 + 41 + 103 + 53 + 18 + 6}{11} = \frac{525}{11} = 47{,}3$$

Lahko trdimo, da napoved porabe zadevnega materiala v trinajsti terminski enoti (tednu, mesecu ...)

z verjetnostjo ~ 0,68 (68%) leži med (235 – 1,25 · 47,3) ≈ 188
in (235 + 1,25 · 47,3) ≈ 312 enotami,

z verjetnostjo ~ 0,95 (95%) leži med (235 – 2,50 · 47,3) ≈ 126
in (235 + 2,50 · 47,3) ≈ 374 enotami,

z verjetnostjo ~ 0,999 (99,9%) leži med (235 – 3,75 · 47,3) ≈ 64
in (235 + 3,75 · 47,3) ≈ 436 enotami.

Ekstrapolacijske metode

Dvojno eksponentno glajenje 1.reda

- Pri napovedovanju pojavov, kjer je mogoče opaziti (linearni) trend, enostavno eksponentno glajenje prvega reda ne daje pravilnih rezultatov; trend namreč smatra kot iregulariteto in ga želi izničiti;
- dvojno eksponentno glajenje prvega reda obravnava posebej osnovno vrednost pojava in posebej vrednost trenda;
- napoved vrednosti pojava v nekem časovnem obdobju je vsota napovedi osnovne vrednosti (baze) in napovedi vrednosti trenda:

$$\boxed{F_i = B_i + T_i}$$

F_i = *napoved vrednosti pojava v obdobju* i, $i = 1 .. n$
B_i = *napoved osnovne vrednosti (baze) pojava v obdobju* i, $i = 1 .. n$
T_i = *napoved vrednosti trenda pojava v obdobju* i, $i = 1 .. n$
n = *število obdobij, za katera imamo podatke.*

Ekstrapolacijske metode
Dvojno eksponentno glajenje 1.reda

- osnovno vrednost pojava se napove po enačbi

$$B_{i+1} = \beta \cdot R_i + (1-\beta) \cdot (B_i + T_i)$$

B_{i+1} = *napoved osnovne vrednosti pojava v obdobju* $i+1$, $i = 1 .. n$
R_i = *dejanska vrednost pojava v obdobju* i, $i = 1 .. n$
B_i = *predhodna napoved osnovne vrednosti pojava v obdobju* i, $i = 1 .. n$
T_i = *predhodna napoved vrednosti trenda pojava v obdobju* i, $i = 1 .. n$
β = *konstanta eksponentnega glajenja osnovne vrednosti pojava,* $0 \leq \beta \leq 1$
n = *število obdobij, za katera imamo podatke*

- vrednost trenda pojava pa se napove po enačbi

$$T_{i+1} = \gamma \cdot (B_{i+1} - B_i) + (1-\gamma) \cdot T_i$$

T_{i+1} = *napoved vrednosti trenda pojava v obdobju* $i+1$, $i = 1 .. n$
γ = *konstanta eksponentnega glajenja vrednosti trenda pojava,* $0 \leq \gamma \leq 1$

- β je običajno med $0{,}2$ in $0{,}5$, γ pa med $0{,}1$ in $0{,}5$.

PRIMER NAPOVEDOVANJA Z METODO DVOJNEGA EKSPONENTNEGA GLAJENJA PRVEGA REDA

Prodajo nekega izdelka, ki v devetih mesecih izkazuje rastoč trend, napovedujemo z dvojnim eksponentnim glajenjem prvega reda s konstanto eksponentnega glajenja osnovne vrednosti pojava β = 0,2 in konstanto eksponentnega glajenja vrednosti trenda pojava γ = 0,4. Za prvi mesec mora biti napoved ocenjena: osnovna vrednost 110 enot in trenda 20 enot, skupno torej 130 enot.

$$B_2 = \beta \cdot R_1 + (1-\beta)\cdot(B_1+T_1) = = 0{,}2\cdot 120 + (1-0{,}2)\cdot(110+20) \approx 128$$

$$T_2 = \gamma\cdot(B_2-B_1) + (1-\gamma)\cdot T_1 = = 0{,}4\cdot(128-110) + (1-0{,}4)\cdot 20 \approx 19$$

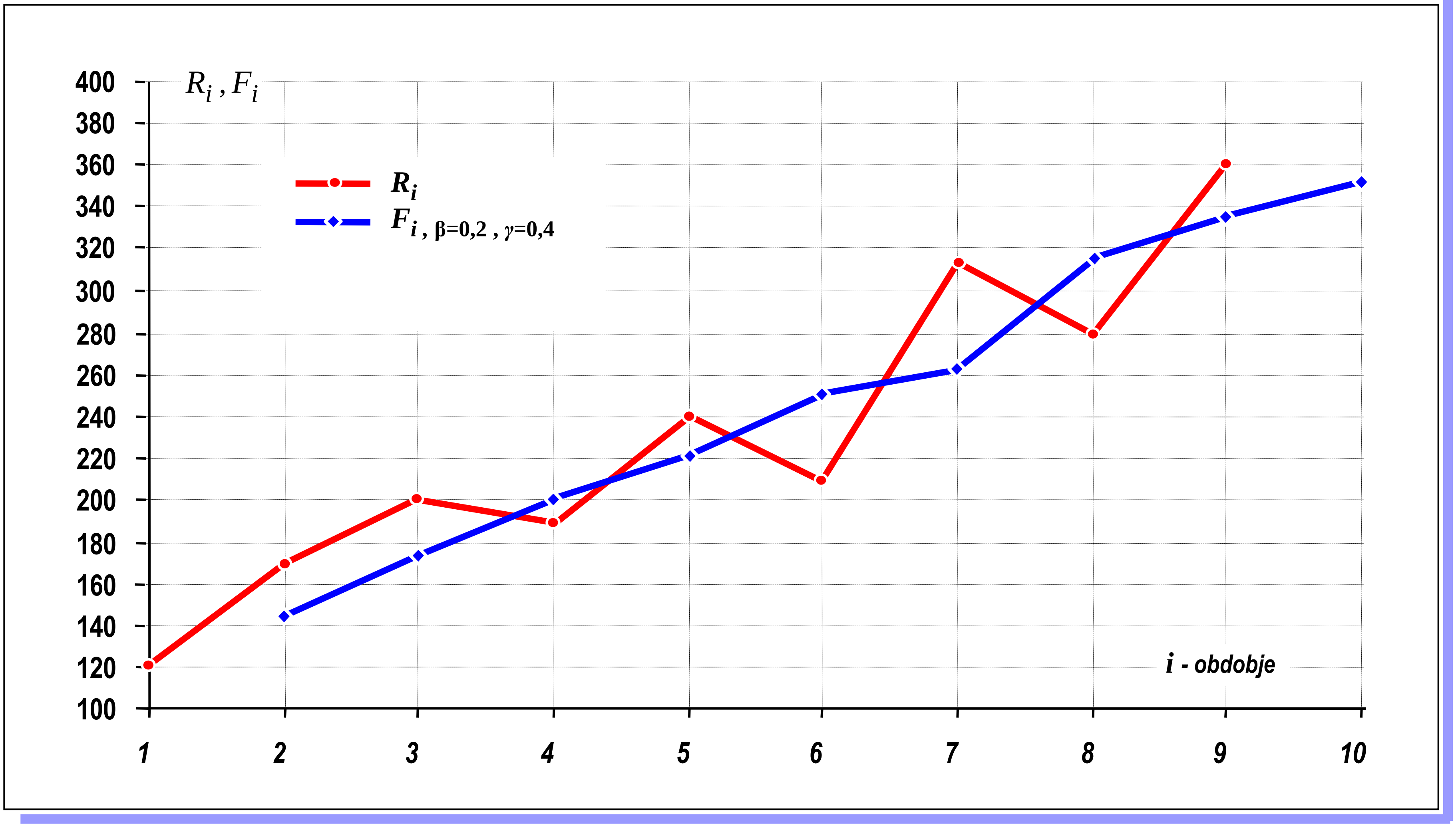
$$F_2 = B_2+T_2 = 128 + 19 = 147$$

$$B_3 = \beta\cdot R_2 + (1-\beta)\cdot(B_2+T_2) = = 0{,}2\cdot 170 + (1-0{,}2)\cdot(128+19) \approx 152$$

$$T_3 = \gamma\cdot(B_3-B_2) + (1-\gamma)\cdot T_2 = = 0{,}4\cdot(152-128) + (1-0{,}4)\cdot 19 \approx 21$$

$$F_3 = B_3+T_3 = 152 + 21 = 173$$

i	1	2	3	4	5	6	7	8	9	10
R_i	120	170	200	190	240	210	310	280	360	
$B_{i,\beta=0,2}$	110	128	152	178	199	225	241	271	293	325
$T_{i,\gamma=0,4}$	20	19	21	23	22	24	21	25	23	27
$F_{i,\beta=0,2,\gamma=0,4}$	130	147	173	201	221	249	262	296	316	352



R_i , F_i
R_i
F_{i} , β=0,2 , γ=0,4
400
380
360
340
320
300
280
260
240
220
200
180
160
140
120
100
i - obdobje
1
2
3
4
5
6
7
8
9
10

Absolutna srednja napaka napovedi (MAD):

i	1	2	3	4	5	6	7	8	9	10
R_i	120	170	200	190	240	210	310	280	360	
F_i		147	173	201	221	249	262	296	316	352
$\lvert R_i - F_i \rvert$		23	27	11	19	39	48	16	16	

Število statističnih podatkov = 8

$$MAD = \frac{23 + 27 + 11 + 19 + 39 + 48 + 16 + 16}{8} = \frac{189}{8} = 23{,}6$$

Lahko trdimo, da napoved prodaje zadevnega izdelka v deseti terminski enoti (mesecu ...)

z verjetnostjo ~ 0,68 (68%) leži med (352 – 1,25 · 23,6) ≈ 323
in (352 + 1,25 · 23,6) ≈ 382 enotami,

z verjetnostjo ~ 0,95 (95%) leži med (352 – 2,50 · 23,6) ≈ 293
in (352 + 2,50 · 23,6) ≈ 411 enotami,

z verjetnostjo ~ 0,999 (99,9%) leži med (352 – 3,75 · 23,6) ≈ 264
in (352 + 3,75 · 23,6) ≈ 441 enotami.

Ekstrapolacijske metode

Eksponentno glajenje višjih redov

- Teoretično je število redov eksponentnega glajenja neskončno, v praksi pa je pomembno eksponentno glajenje drugega reda:

$$F''_{i+1} = F''_i + \alpha'' \cdot \left(F'_{i+1} - F''_i \right)$$

oziroma

$$F''_{i+1} = (1-\alpha'') \cdot F''_i + \alpha'' \cdot F'_{i+1}$$

F''_{i+1} = *napoved drugega reda vrednosti pojava v obdobju* $i+1$, $i = 1 .. n$,

F''_i = *predhodna napoved drugega reda vrednosti pojava v obdobju* i, $i = 1 .. n$,

F'_{i+1} = *napoved prvega reda vrednosti trenda pojava v obdobju* $i+1$, $i = 1 .. n$,

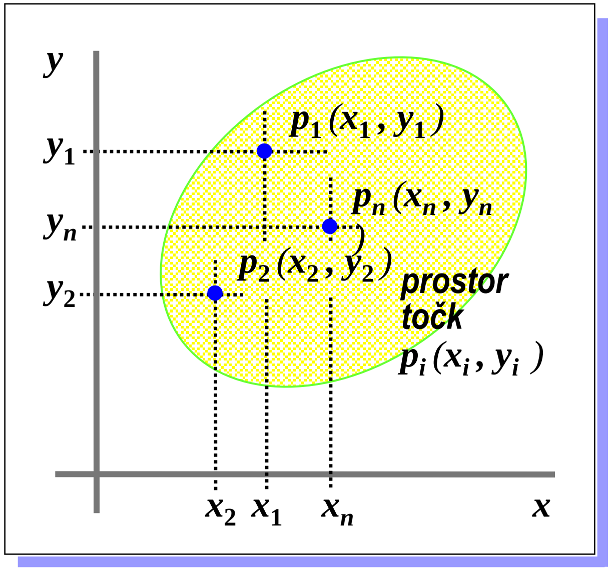
α'' = *konstanta (faktor) eksponentnega glajenja drugega reda,* $0 \le \alpha'' \le 1$

n = *število obdobij, za katera imamo podatke*

- Uporabno je za napovedovanje pojavov z nelinearnim trendom;
- konstanta glajenja α" je običajno okrog 0,1.

Korelacijske metode napovedovanja

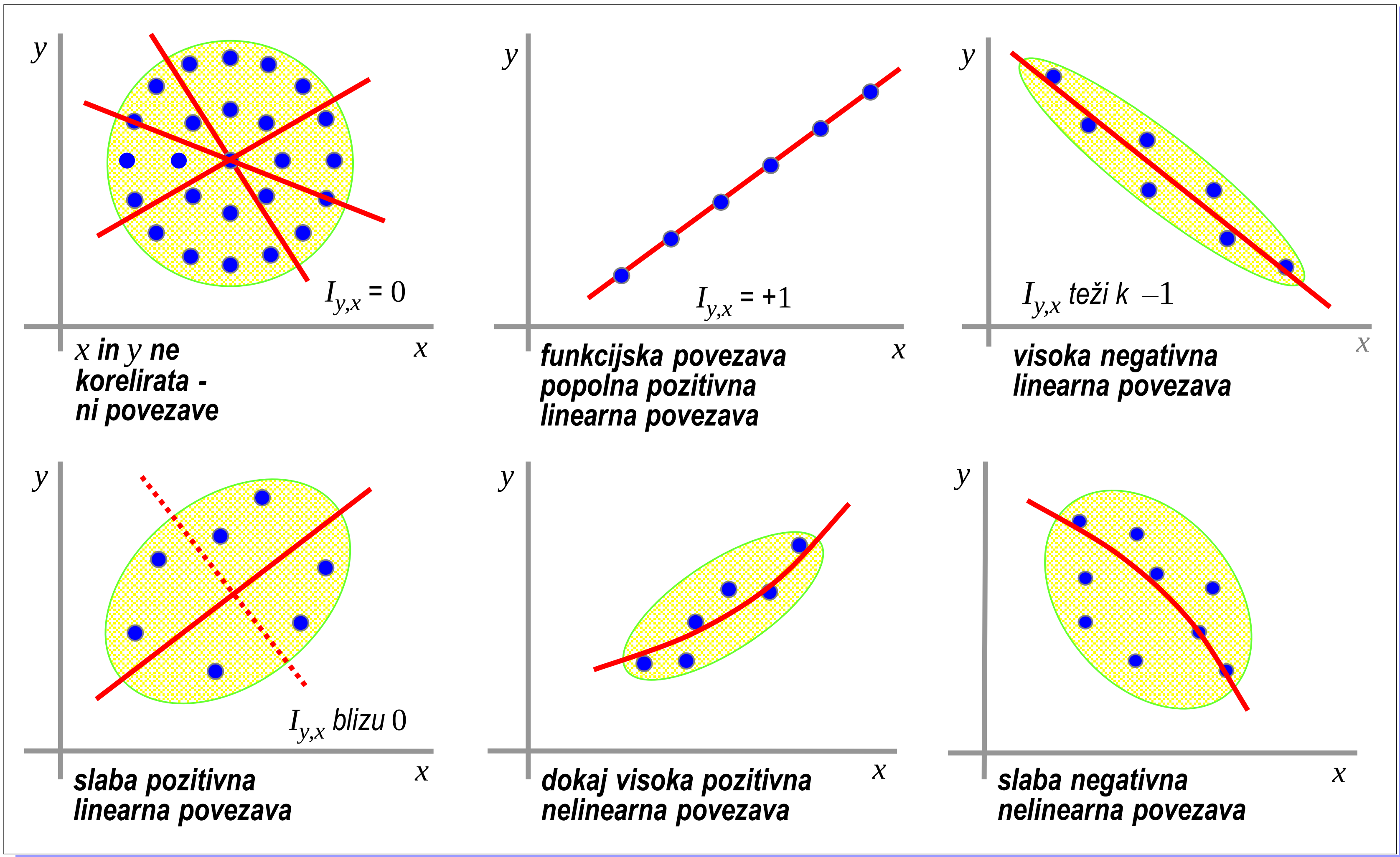
- Procesi ali pojavi, ki so odvisni le sami od sebe in od ničesar drugega, so redki,
- običajno <u>so pojavi medsebojno povezani in soodvisni</u>, en pojav zavisi od drugega,
- vsaki vrednosti enega pojava (neodvisne spremenljivke - vzroka, x) ustreza neka vrednost drugega pojava (odvisne spremenljivke - posledice, y), kar lahko predstavimo v ravninskem koordinatnem sistemu;



- <u>Proučevanje povezav med pojavi</u>:
  - <u>korelacija</u>: ali obstoji povezava med dvema pojavoma? kako močna je ta povezava?
  - <u>regresija</u>: kakšna je oblika te povezave?

Korelacijske metode

- Oblike korelacijskih povezav

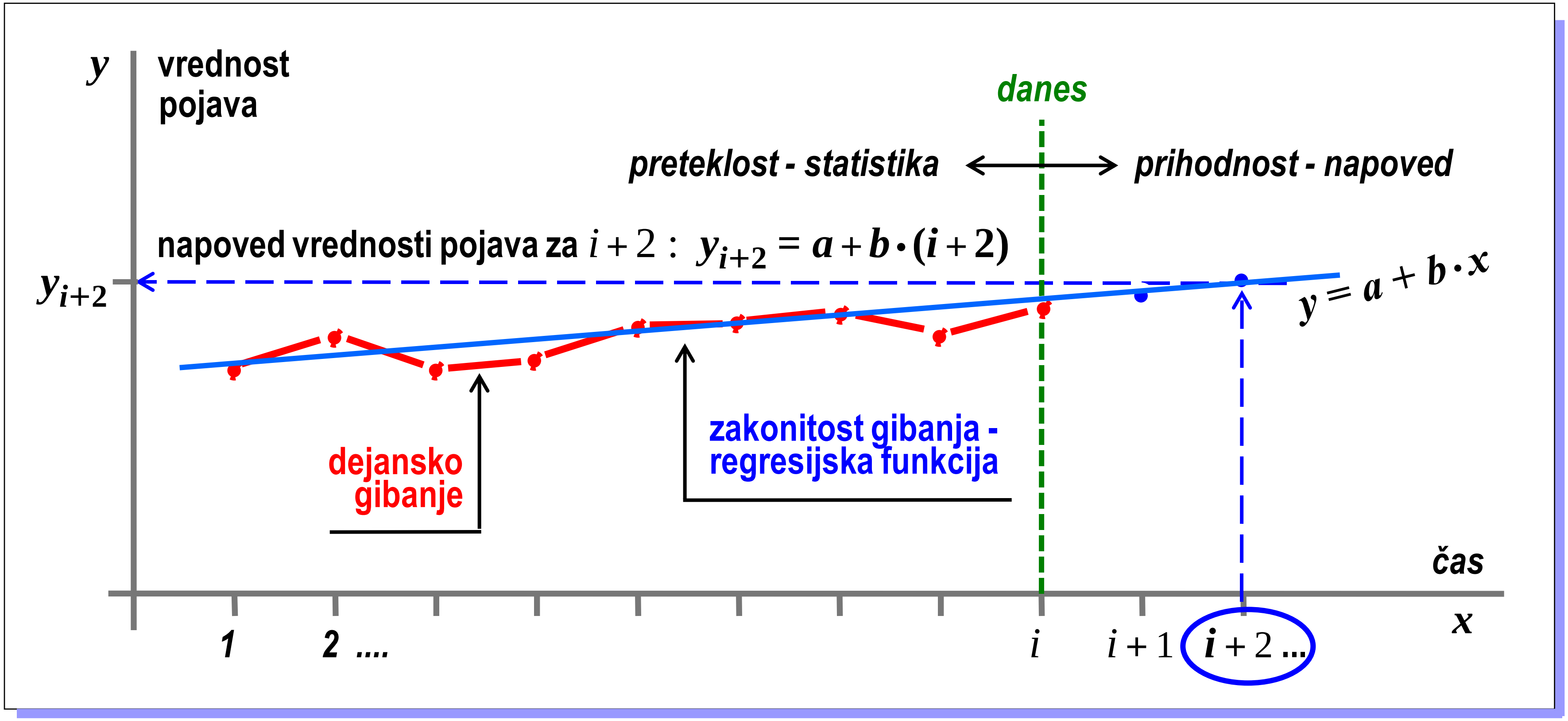


Korelacijske metode

- Obnašanje regresije kaže koeficient korelacije,
- Koeficient korelacije ima vrednost med -1 in $+1$, predznak kaže smer povezave (negativno - padajoče, pozitivno - rastoče)
- za velike statistične vzorce je kvadrat koeficienta korelacije (r^2) koeficient determinacije;
- le-ta leži med 0 in 1 in pove, kolikšen del skupne variance je pojasnjen s povezavo med x in y,
- čim večji je r^2, toliko bolj ustrezno regresijska funkcija ponazarja gibanje pojava.

Korelacijske metode

- Napoved vrednosti pojava (y) za naslednja obdobja izvedemo tako, da po določeni regresijski funkciji povečujemo vrednost neodvisne spremenljivke (x) in izračunavamo pripadajočo vrednost (y).



Korelacijske metode

Linearna regresija 1.reda

- Funkcija, ki se najbolj prilega pojavu, je premica z enačbo $\boxed{y = a + b \cdot x}$
  - Regresijska konstanta a je osnovna vrednost pojava ob (imaginarnem) času nič, regresijski koeficient b pa linearni trend - prirastek ali zmanjšanje vrednosti pojava v enem časovnem obdobju;
  - koeficient determinacije (določenosti) r^2 oceni, ali je regresijska funkcija res blizu premice; leži med 0 in 1; r^2 naj bo > $0{,}75$, če naj premica ustrezno ponazori pojav;
  - koeficient korelacije r ima isti predznak, kot regresijski koeficient b, njegova vrednost pa je med -1 in $+1$; $+1$ pomeni, da gre za popolno pozitivno povezavo, -1 pa, da gre za popolno negativno povezavo

Korelacijske metode

Linearna regresija 1.reda

- Izračun po metodi najmanjših kvadratov

$$a \cdot n + b \cdot \sum x = \sum y$$
$$a \cdot \sum x + b \cdot \sum (x^2) = \sum x \cdot y$$

$$b = \frac{n \cdot \sum x \cdot y - \sum x \cdot \sum y}{n \cdot \sum (x^2) - (\sum x)^2}$$

$$a = \frac{\sum y - b \cdot \sum x}{n}$$

normalne enačbe

a = *regresijska konstanta*
b = *regresijski koeficient*
x = *neodvisna spremenljivka – čas, štejemo ga od 1 naprej*
y = *odvisna spremenljivka – opazovani pojav*
n = *število obdobij, za katera imamo podatke*

$$r^2 = \frac{\left[\sum x \cdot y - \frac{1}{n} \cdot \sum x \cdot \sum y\right]^2}{\left[\sum (x^2) - \frac{1}{n} \cdot (\sum x)^2\right] \cdot \left[\sum (y^2) - \frac{1}{n} \cdot (\sum y)^2\right]}$$

r^2 = *koeficient determinacije (določenosti),* $r = \sqrt{r^2}$ = *indeks korelacije*

PRIMER NAPOVEDOVANJA Z METODO LINEARNE REGRESIJE PRVEGA REDA

Potrebe po nekem materialu, za katere imamo statistične podatke za preteklih 10 mesecev, želimo za naslednja dva meseca napovedati z linearno regresijo prvega reda. Izračun regresije izvedemo po metodi najmanjših kvadratov; izračunane podatke prikažemo tabelarično in grafično. Po potrebi podatke skaliramo s faktorjem 1000. Ustreznost regresijske funkcije preverjamo z indeksom determinacije in koeficientom korelacije, območje napovedi s povprečno napako napovedi.

x - mesec	1	2	3	4	5	6	7	8	9	10
y - prodaja	4012	4296	4519	4431	4594	4878	4992	5327	5561	5720

'Ročni' izračun linearne regresije prvega reda po metodi najmanjših kvadratov:

x	R_i	y	x^2	y^2	$x \cdot y$
1	4012	4,0	1	16,0	4,0
2	4296	4,3	4	18,5	8,6
3	4519	4,5	9	20,3	13,5
4	4431	4,4	16	19,4	15,6
5	4594	4,6	25	21,2	23,0
6	4878	4,9	36	24,0	29,4
7	4992	5,0	49	25,0	35,0
8	5327	5,3	64	28,1	42,4
9	5561	5,6	81	31,4	50,4
10	5720	5,7	100	32,5	55,0
$\sum x$ 55		$\sum y$ 48,3	$\sum(x^2)$ 385	$\sum(y^2)$ 236,2	$\sum x \cdot y$ 280,9
$(\sum x)^2$ 3025		$(\sum y)^2$ 2333			n 10

Izračun linearne regresije prvega reda po metodi najmanjših kvadratov:

Regresijska konstanta in koeficient:

$$b = \frac{n \cdot \sum x \cdot y - \sum x \cdot \sum y}{n \cdot \sum (x^2) - (\sum x)^2} = \frac{10 \cdot 280{,}9 - 55 \cdot 48{,}3}{10 \cdot 385 - 3025} = 0{,}185$$

$$a = \frac{\sum y \; - \; b \; \cdot \sum x}{n} = \frac{48{,}3 - 0{,}185 \cdot 55}{10} = 3{,}813$$

Regresijska funkcija ima torej obliko: $y = a + b \cdot x = 3{,}813 + 0{,}185\, x$.

Koeficienta determinacije in korelacije:

$$r^2 = \frac{\left[\sum x \cdot y \; - \; \frac{1}{n} \; \cdot \sum x \cdot \sum y\right]^2}{\left[\sum (x^2) \; - \; \frac{1}{n} \cdot (\sum x)^2\right] \cdot \left[\sum (y^2) \; - \; \frac{1}{n} \cdot (\sum y)^2\right]} =$$

$$= \frac{[280{,}9 - 0{,}1 \cdot 55 \cdot 48{,}3]^2}{[385 - 0{,}1 \cdot 3025] \cdot [236{,}2 - 0{,}1 \cdot 2333]} = 0{,}965 \qquad r = \sqrt{r^2} = \sqrt{0{,}965} = 0{,}982$$

Izračun povprečne absolutne napake napovedi - MAD:

V regresijsko funkcijo vstavljamo x = 1 do 10 in ugotovimo navidezno napoved za ustrezne terminske enote, nato ugotavljamo absolutno vrednost razlike do dejanske vrednosti potrebe:

$$
\begin{array}{llll}
x = 1 & y_1 = 3{,}813 + 0{,}185 \cdot 1 = 3{,}998 & |R_1 - F_1| = 4012 - 3998 = & 14 \\
x = 2 & y_2 = 3{,}813 + 0{,}185 \cdot 2 = 4{,}173 & |R_2 - F_2| = 4296 - 4173 = & 123 \\
x = 3 & y_3 = 3{,}813 + 0{,}185 \cdot 3 = 4{,}358 & |R_3 - F_3| = 4519 - 4358 = & 161 \\
x = 4 & y_4 = 3{,}813 + 0{,}185 \cdot 4 = 4{,}543 & |R_4 - F_4| = 4431 - 4543 = & 112 \\
x = 5 & y_5 = 3{,}813 + 0{,}185 \cdot 5 = 4{,}728 & |R_5 - F_5| = 4594 - 4728 = & 134 \\
x = 6 & y_6 = 3{,}813 + 0{,}185 \cdot 6 = 4{,}913 & |R_6 - F_6| = 4878 - 4913 = & 35 \\
x = 7 & y_7 = 3{,}813 + 0{,}185 \cdot 7 = 5{,}098 & |R_7 - F_7| = 4992 - 5098 = & 106 \\
x = 8 & y_8 = 3{,}813 + 0{,}185 \cdot 8 = 5{,}283 & |R_8 - F_8| = 5327 - 5283 = & 44 \\
x = 9 & y_9 = 3{,}813 + 0{,}185 \cdot 9 = 5{,}468 & |R_9 - F_9| = 5561 - 5468 = & 93 \\
x = 10 & y_{10} = 3{,}813 + 0{,}185 \cdot 10 = 5{,}653 & |R_{10} - F_{10}| = 5720 - 5653 = & 67 \\
\hline
 & & & = 822
\end{array}
$$

$$MAD = \frac{\sum_{i=1}^{n} \left| (R_i - F_i) \right|}{n} = \frac{822}{10} = 82{,}2$$

PRIMER NAPOVEDOVANJA Z METODO LINEARNE REGRESIJE PRVEGA REDA

x - *mesec*	1	2	3	4	5	6	7	8	9	10
y - prodaja	4012	4296	4519	4431	4594	4878	4992	5327	5561	5720

Regresijska funkcija: $y = a + b \cdot x = 3{,}813 + 0{,}185 \cdot x$

Koeficient korelacije: $r^2 = 0{,}965$ →

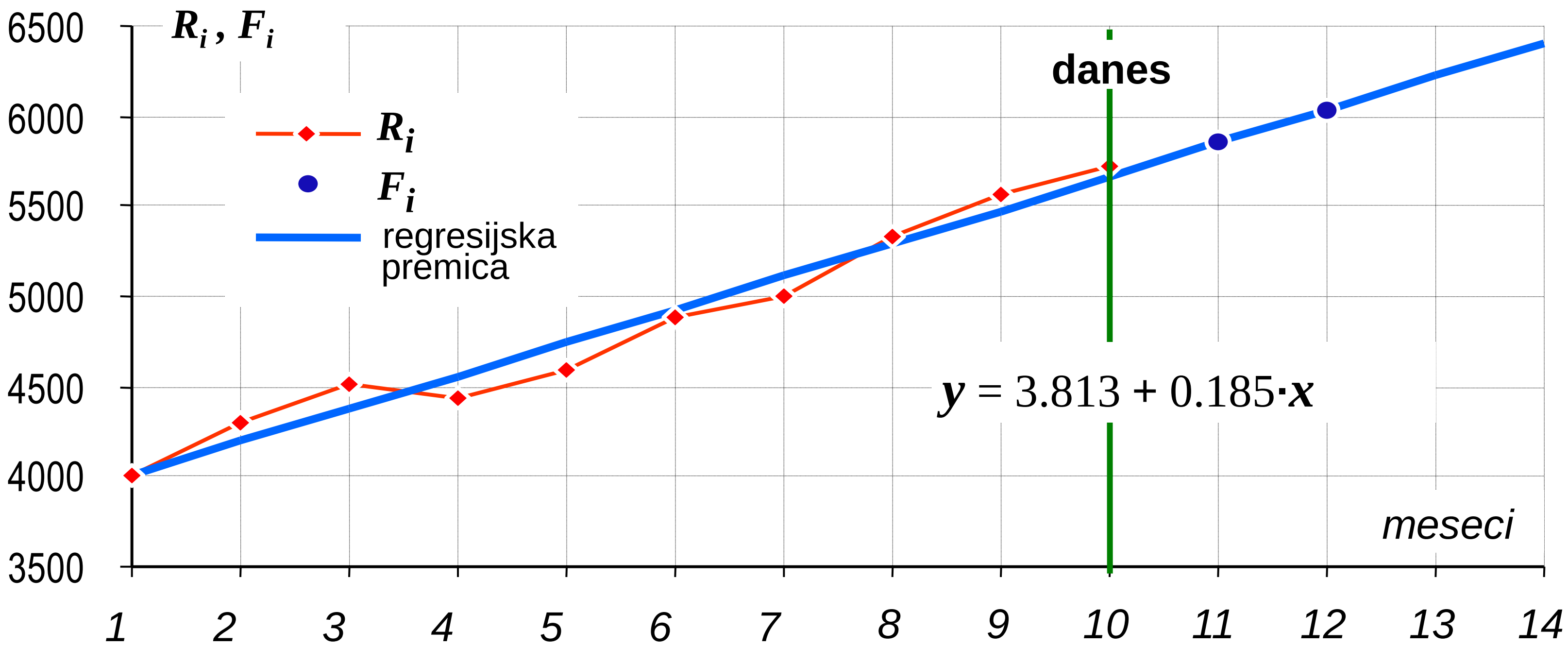
premica dovolj dobro ponazarja gibanje pojava - potreb

Indeks determinacije: $r = 0{,}982$ →

potrebe so odvisne od časa

Srednja napaka napovedi: $MAD = 0{,}00822$ →

z verjetnostjo 0,95 bo napoved med izračunano vrednostjo minus 0,02055 in izračunano vrednostjo plus 0,02055



Napoved za enajsti in dvanajsti mesec:

$x = 11 \rightarrow y_{11}$ = 3.813 + 11· 0,185 = 5,848 $\rightarrow F_{11}$ = 1000· 5,848 = 5848

z verjetnostjo 0,95 lahko trdimo, da bodo potrebe v enajstem mesecu med F_{11} – 2,5 MAD in F_{11} + 2,5 MAD, torej med 5643 in 6063 enotami;

$x = 12 \rightarrow y_{12}$ = 3.813 + 12· 0,185 = 6,033 $\rightarrow F_{12}$ = 1000· 6,033 = 6033

z verjetnostjo 0,95 lahko trdimo, da bodo potrebe v dvanajstem mesecu med 5828 in 6238 enotami.

Korelacijske metode

Nelinearna regresija 1.reda

- Premica kot regresijska funkcija se opazovanemu pojavu vedno ne prilega najbolje; morda je neka krivulja za ponazoritev zakonitosti pojava primernejša;
- krivulje prvega reda (x ima eksponent 1!), ki jih srečujemo pri tem:
  - eksponentna funkcija
  - logaritemska funkcija I.
  - logaritemska funkcija II.
  - logaritemska funkcija III.

$$y = a \cdot b^x \rightarrow$$
$$\log y = \log a + x \cdot \log \mathrm{b}$$
$$y = a + b \cdot \log x$$
$$\log y = a + b \cdot x$$
$$\log y = a + b \cdot \log x$$

- Izračunava se jih lahko prav tako po metodi najmanjših kvadratov; ne sme se pozabiti na morebitno antilogaritmiranje!

Korelacijske metode

Regresije višjih redov

- so podobne nelinearni regresiji prvega reda z razliko, da imamo opravka z regresijskimi krivuljami višjih redov, ki pa ne smejo biti ciklične,
- praktičen pomen ima regresija drugega reda z regresijskimi funkcijami:
  - $y = a + b \cdot x + c \cdot x^2$
  - $y = a + b \cdot x + c \cdot \sqrt{x}$
  - $y = a \cdot b^x \cdot c^{x^2}$

    oziroma $\log y = \log a + x \cdot \log b + x^2 \cdot \log c$

- Tudi regresije višjih redov se izračunava iz sistemov normalnih enačb po metodi najmanjših kvadratov; ne sme se pozabiti na morebitno antilogaritmiranje!

Korelacijske metode

Večstopenjska (multipla) regresija

- Do sedaj smo predpostavljali, da je pojav, ki ga obravnavamo, odvisen samo od časa oziroma od neke druge neodvisne spremenljivke;
- kadar pa je pojav poleg odvisnosti od časa dodatno povezan še z drugimi dejavniki (npr. prodaja neke vrste izdelka je lahko odvisna od kupne moči, časa, splošnega trenda gospodarske rasti, reklame ipd.), je treba upoštevati kompleksno odvisnost pojava od večih faktorjev (in le eden med njimi je morda tudi čas);
- če se ugotovi vplive vsakega posameznega dejavnika, se jih izloči in prikaže v posebni regresijski funkciji, se lahko tudi predvidi, kako bo vsak od njih vplival na gibanje pojava v prihodnosti oziroma kakšno bo gibanje pojava pod skupnim vplivom vsem dejavnikov.

Korelacijske metode

Večstopenjska - multipla regresija

- Stohastično povezavo - korelacijo med pojavom in dejavniki, ki vplivajo nanj, se v takem primeru izrazi v obliki

$$\boxed{y = \mathrm{f}(x_1, x_2, .., x_n) + \varepsilon}$$

- Regresijska funkcija je sedaj ploskev v nekem splošnem večdimenzionalnem prostoru;
- večstopenjska regresijska analiza ima dve nalogi:
  - najprej določitev faktorjev, ki imajo poleg časa dejanski vpliv na opazovani pojav; pri čemer morajo le-ti biti statistično dokazani,
  - in nato določitev zakonitosti pojava v obliki funkcije, ki prikazuje relacije med posameznimi spremenljivkami; enačba regresijske funkcije sme vsebovati le tiste neodvisne spremenljivke, katerih vpliv je statistično dokazan.

Proces napovedovanja