

Curs 10

Analiza și prognoza seriilor temporale (I)

Modelarea și analiza datelor pentru decizii de management

Conf.dr.ing. Lăcrimioara GRAMA

Anul 2 master · Tehnologii, Sisteme și Aplicații pentru eActivități

Facultatea de Electronică, Telecomunicații și Tehnologia Informației
Universitatea Tehnică din Cluj-Napoca



Cursul precedent

Analiza decizională (II)

- ✓ Analiza deciziilor cu eşantion de informații
 - ✓ Diagrama de influență
 - ✓ Arborele și strategia de decizie
 - ✓ Profilul de risc
 - ✓ Valoarea aşteptată și eficiența eşantionului de informații
- ✓ Actualizarea probabilităților
 - ✓ Probabilități preliminare, condiționate și posterioare
 - ✓ Aplicarea teoremei lui Bayes
- ✓ Studii de caz
 - ✓ Bayer Pharmaceuticals
 - ✓ Centrul Medical al Universității Duke
 - ✓ Oglethorpe Power



La finalul acestui curs, studenții vor putea să:

- diferențieze metodele calitative și cantitative de prognoză
- identifice tiparele orizontal, de trend, sezonier, de trend și sezonaliitate și ciclic
- calculeze și interpreteze erorile de prognoză și indicatorii ME, MAE, MSE și MAPE
- aplice metoda naivă, media valorilor trecute, media alunecătoare și media alunecătoare ponderată
- aplice netezirea exponențială simplă și să interpreteze influența constantei de netezire α
- compare și selecteze metodele de prognoză în funcție de acuratețe și de tiparul seriei temporale

Obiectivele cursului



Cuprins

Analiza previzională

- Metode cantitative de prognoză

Tipare ale seriilor temporale

- Tipar orizontal
- Tipar de trend
- Tipar de sezonalitate / sezonier
- Tipar de trend și sezonalitate
- Tipar ciclic
- Selectarea unei metode de prognoză

Acuratețea prognozei

Metode de prognoză pentru serii cu tipar orizontal

- Media alunecătoare
- Media alunecătoare ponderată
- Netezirea exponențială

Rezumat

Analiza previzională

Planificarea este unul dintre rolurile esențiale ale managementului

Succesul pe termen lung al unei organizații depinde de capacitatea acesteia de:

- a-și stabili obiective relevante
- a anticipa evoluțiile, oportunitățile și riscurile
- a transforma obiectivele în decizii și acțiuni



Anticiparea viitorului – termeni utilizați

- **Prospectiva / anticiparea strategică** (*foresight*) explorează evoluții și scenarii viitoare posibile, pentru fundamentarea deciziilor pe termen lung
- **Predicția** (*prediction*) estimează un rezultat sau un eveniment viitor pe baza datelor, modelelor, experienței sau raționamentului expert
- **Prognoza** (*forecasting*) produce estimări pentru un orizont de timp definit, prin metode calitative sau cantitative

Metode calitative și cantitative de prognoză

Metodele calitative

Se bazează în principal pe **judicata și experiența experților**

Sunt adecvate când:

- datele istorice nu sunt disponibile sau sunt insuficiente
- condițiile viitoare diferă semnificativ de cele din trecut
- decizia implică informații dificil de cuantificat

Metodele cantitative

Utilizează **date istorice și modele matematice sau statistice**

Pot fi aplicate când:

- sunt disponibile date istorice relevante
- informațiile pot fi cuantificate
- este rezonabil să presupunem că tiparele sau relațiile observate vor continua în orizontul de prognoză



În practică, metodele **calitative** și **cantitative** sunt adesea **combinate**



Metode cantitative de prognoză

Metodele cantitative de prognoză folosesc date istorice și modele matematice sau statistice pentru **estimarea valorilor viitoare**

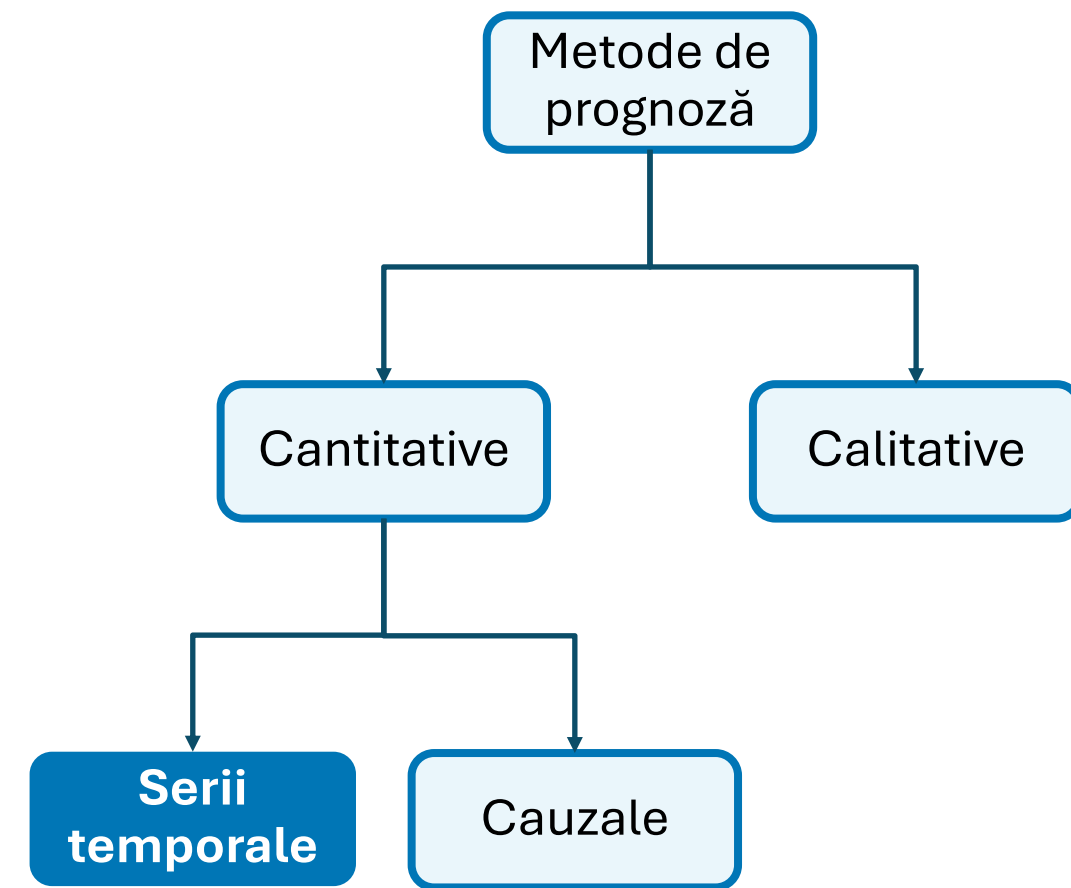
O serie temporală (*time series*) este o succesiune de observații ale aceleiași variabile, ordonate cronologic și, de regulă, înregistrate la intervale egale

Metode bazate pe serii temporale

Utilizează valorile anterioare ale variabilei prognozate și, în unele modele, **erorile de prognoză anterioare**

Metode cauzale

Utilizează una sau mai multe **variabile explicative** pentru care se presupune că influențează variabila prognozată



Tema acestui curs

Cuprins

Analiza previzională

Tipare ale seriilor temporale

- Tipar orizontal
- Tipar de trend
- Tipar de sezonabilitate / sezonier
- Tipar de trend și sezonabilitate
- Tipar ciclic
- Selectarea unei metode de prognoză

Acuratețea prognozei

Metode de prognoză pentru serii cu tipar orizontal

Rezumat

Tiparele seriilor temporale



Obiectiv

Identificarea unui tipar în datele istorice și extrapolarea acestuia pentru perioade viitoare

- Prognoza se bazează pe valorile anterioare ale variabilei și, în anumite modele, pe erorile de prognoză anterioare

Seria temporală

O **serie temporală** este o succesiune de observații ale aceleiași variabile, ordonate cronologic

- Timpul este **discret** — oră, zi, săptămână, lună, trimestru, an — sau **continuu**
- În acest curs analizăm serii temporale discrete, cu observații consecutive înregistrate la intervale egale

Identificarea tiparului unei serii temporale

Tiparul seriei temporale (*data pattern*) descrie modul în care valorile seriei evoluează în timp

- Dacă există motive să presupunem că tiparul observat va continua, acesta poate ghida selectarea unei metode de prognoză adecvate

Primul pas: reprezentarea grafică

- timpul – pe abscisă (axa orizontală)
- valorile seriei – pe ordonată (axa verticală)

Tipare identificate prin examinarea graficului



Tipar orizontal

Valorile fluctuează aleatoriu în jurul unui nivel mediu aproximativ constant

Ex: Tabelul prezintă vânzările săptămânale de benzină ale unui distribuitor din Cluj-Napoca, pentru 12 săptămâni consecutive.

Săptămâna	1	2	3	4	5	6	7	8	9	10	11	12
Vânzări [mii litri]	59.5	73.5	66.5	80.5	63	56	70	63	77	70	52.5	77

Media observată: **67,375 litri** pe săptămână

- Fluctuațiile din jurul mediei nu indică o tendință sistematică; seria prezintă un **tipar orizontal**

Tipar orizontal și staționaritate

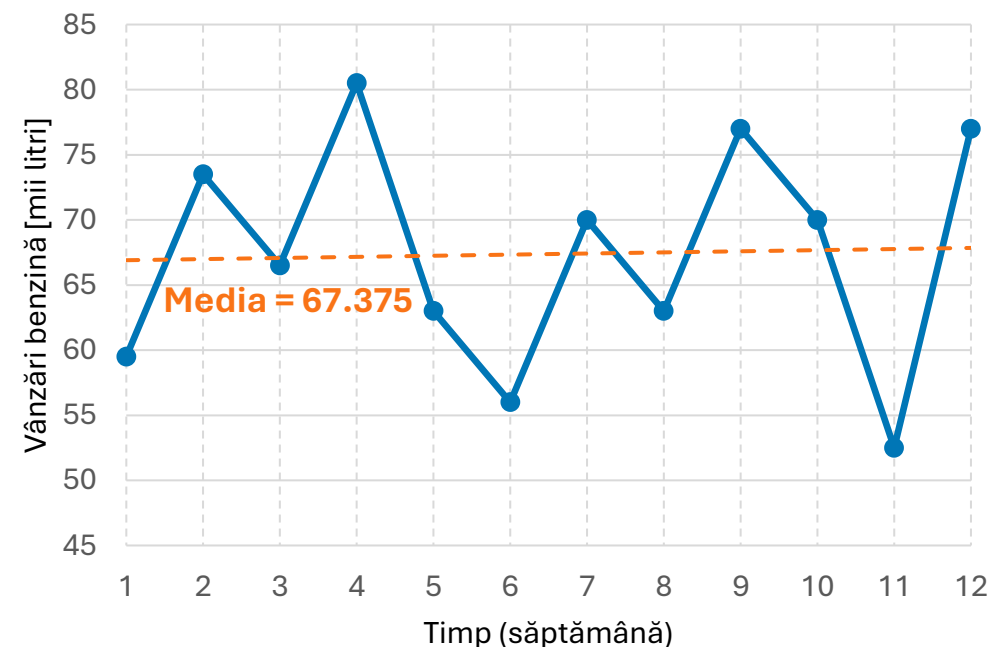
Serie temporală staționară – serie ale cărei proprietăți statistice nu se modifică în timp:

- 1) media este constantă
- 2) variabilitatea este constantă
- 3) covarianța dintre două observații depinde doar de distanța temporală dintre ele, nu de momentul înregistrării

O serie temporală staționară prezintă, de regulă, un tipar orizontal

- Invers, un tipar orizontal **nu este suficient** pentru a demonstra că seria este staționară

Graficul seriei temporale



Tipar orizontal: schimbare de nivel

O modificare a condițiilor de afaceri poate determina trecerea bruscă a seriei temporale la un nou nivel mediu

Săptămâna	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Vânzări [mii litri]	59.5	73.5	66.5	80.5	63	56	70	63	77	70	52.5	77	108.5	119	108.5	115.5	98	112	105	101.5	119	115.5

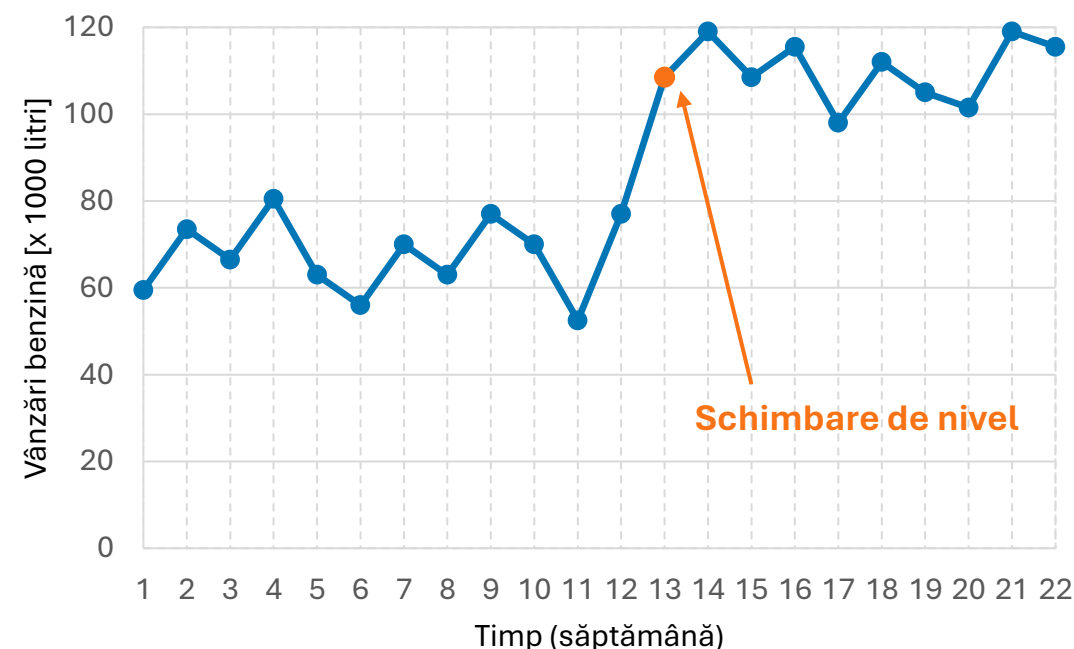
Ex: Începând cu săptămâna 13, distribuitorul încheie un contract cu durata de 10 săptămâni pentru alimentarea vehiculelor Poliției Locale din Cluj-Napoca. Cererea suplimentară determină o creștere bruscă a vânzărilor.

- **Schimbare de nivel în săptămâna 13:** media vânzărilor crește de la aproximativ 67,375 la 110,250 litri pe săptămână

Implicație managerială

Metoda de prognoză trebuie să netezească fluctuațiile aleatorii și să se adapteze rapid la schimbarea nivelului seriei

Graficul seriei temporale



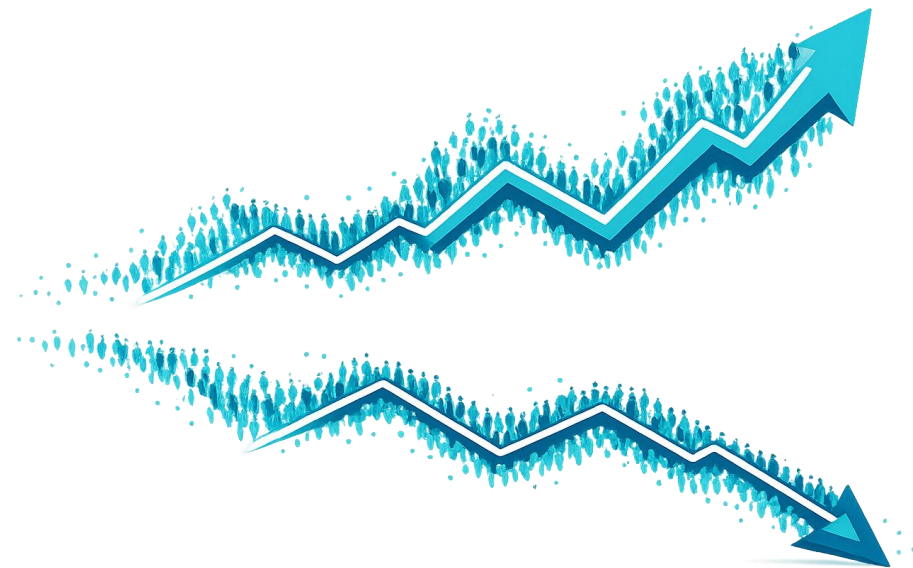
Tipar de trend

Trendul reprezintă direcția generală de evoluție a unei serii temporale pe termen lung

- În pofida fluctuațiilor aleatorii, nivelul seriei tinde să crească sau să scadă treptat
- Trendul poate fi **ascendent** sau **descendent** și poate avea o evoluție **liniară** sau **neliniară**

Factori care pot genera un trend

- creșterea sau scăderea populației
- modificarea structurii demografice
- progresul tehnologic
- schimbarea preferințelor consumatorilor
- schimbările economice sau legislative



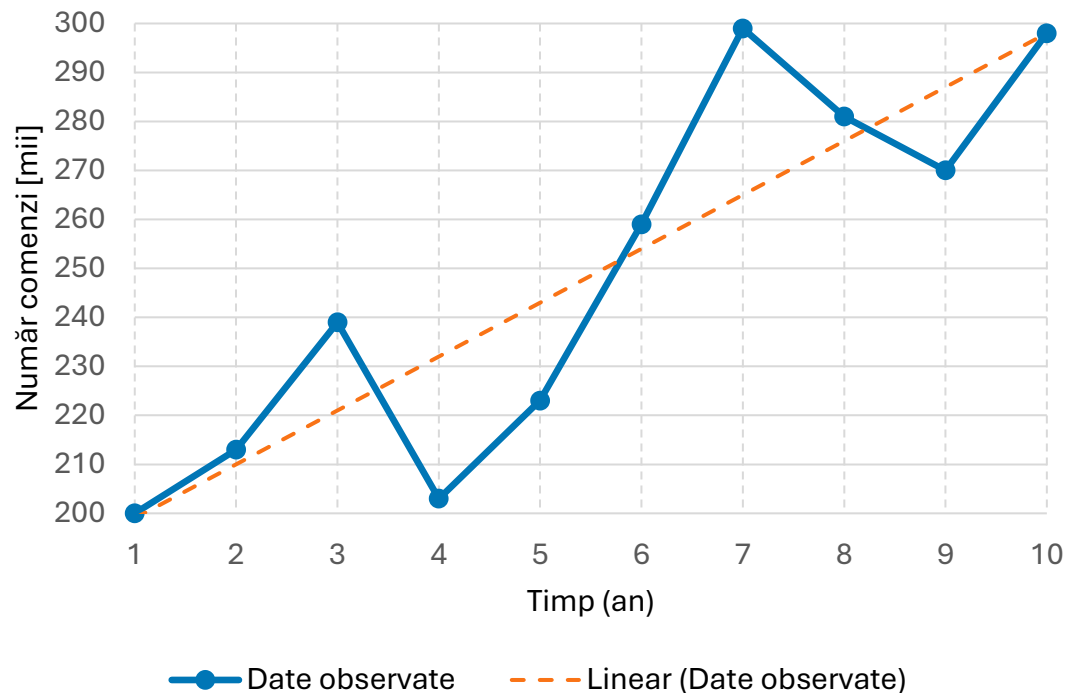
Fluctuațiile de la o perioadă la alta nu exclud existența unui trend; relevantă este direcția generală pe termen lung

Trend liniar ascendent

Ex: Tabelul prezintă numărul anual de comenzi procesate de un magazin online în ultimii 10 ani.

An	1	2	3	4	5	6	7	8	9	10
Număr comenzi [mii]	200	213	239	203	223	259	299	281	270	298

Graficul seriei temporale



Interpretare

- Numărul comenzilor fluctuează de la un an la altul
- Cu toate acestea, direcția generală a seriei este **ascendentă**

Evoluția poate fi aproximată printr-un **trend liniar ascendent**

Fluctuațiile temporare nu contrazic existența unui trend, dacă direcția generală a seriei rămâne clară

Trend neliniar ascendent

Unele serii temporale nu pot fi approximate adecvat printr-un trend liniar. În aceste situații, se utilizează modele neliniare care descriu mai bine evoluția datelor

Ex: Tabelul prezintă evoluția anuală a numărului de utilizatori activi ai unei platforme digitale globale, pe parcursul a 10 ani.

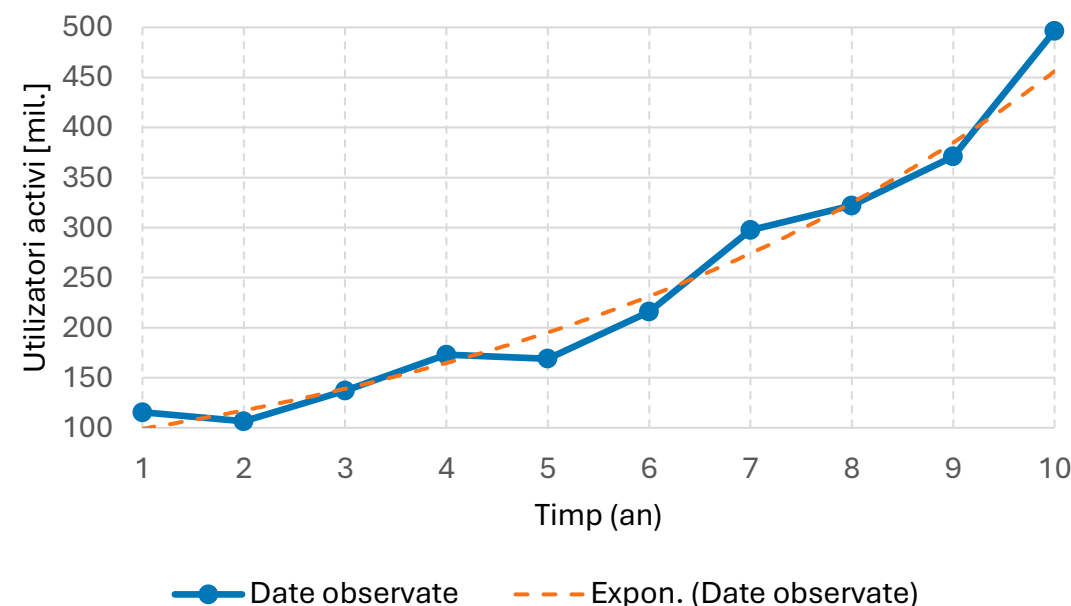
An	1	2	3	4	5	6	7	8	9	10
Utilizatori activi [mil.]	115.5	106.5	137	173	169	216	297.5	322	371	496.5

Interpretare

- Seria prezintă un **trend ascendent neliniar**
- Numărul utilizatorilor nu crește cu aceeași valoare absolută de la un an la altul
- Pe ansamblu, ritmul creșterii se accentuează, sugerând o evoluție aproximativ exponențială

Un **model exponențial** este adecvat atunci când **rata procentuală de variație** de la o perioadă la alta este **aproximativ constantă**

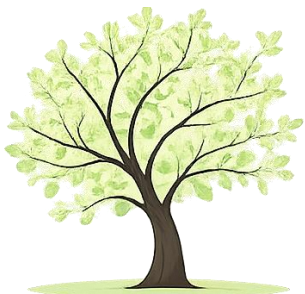
Graficul seriei temporale



Tipar de sezonabilitate

Tiparul de sezonabilitate (*seasonal pattern*) reprezintă un tipar care se repetă la intervale regulate, ca efect al unor influențe calendaristice, climatice sau comportamentale

- Perioada de repetare este, de regulă, **de cel mult un an**
- Tiparul poate apărea **lunar, trimestrial, săptămânal, zilnic** sau chiar **în interiorul aceleiași zile**



Mesaj-cheie

Sezonabilitatea este dată de **regularitatea repetării** unui comportament, nu de lungimea totală a seriei

- **Sezonabilitate anuală** – determinată de anotimpuri, sărbători etc.
- **Sezonabilitate subanuală** – determinată de zilele săptămânii, orele zilei etc.

Exemple

- Vânzările de piscine cresc primăvara și vara și scad toamna și iarna
- Vânzările de echipamente pentru sporturile de iarnă urmează, de regulă, tiparul opus
- Traficul rutier prezintă un tipar zilnic, cu valori maxime în orele de vârf

Tipar de sezonabilitate: exemplu

Ex: Tabelul prezintă vânzările trimestriale de produse pentru grădinărit ale unui magazin, pe parcursul a 5 ani.

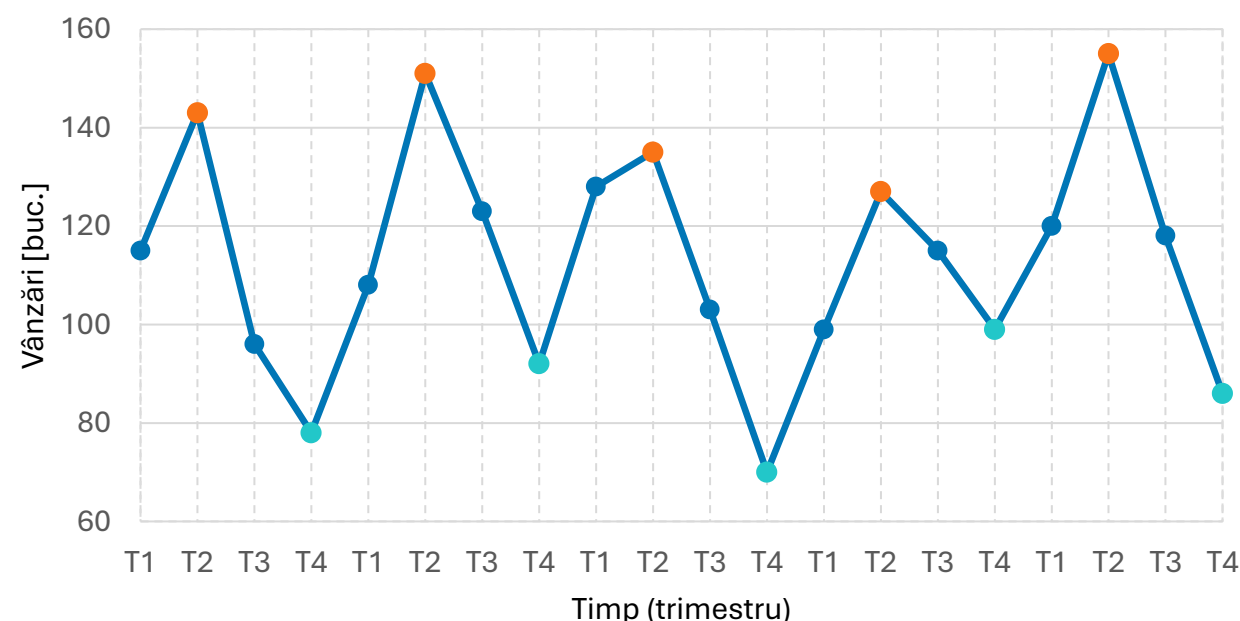
An	1				2				3				4				5			
Trimestru	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Vânzări [buc.]	115	143	96	78	108	151	123	92	128	135	103	70	99	127	115	99	120	155	118	86

Interpretare

- Seria nu prezintă o tendință clară de creștere sau de scădere pe termen lung
- Tiparul se repetă la fiecare **4 trimestre**
- Vânzările sunt, în general, **maxime în trimestrul 2** și **minime în trimestrul 4**

Fluctuațiile pot masca sezonabilitatea; tiparul devine evident prin compararea aceluiași trimestre din ani succesivi

Graficul seriei temporale



Tipar combinat: trend și sezonaliitate

Unele serii temporale combină o **evoluție pe termen lung** cu variații sezoniere care se repetă la intervale regulate — **tipar de trend și sezonaliitate** (*trend and seasonal pattern*)

Ex: Tabelul prezintă vânzările trimestriale de tablete ale unui magazin pe parcursul a 4 ani.

An	1				2				3				4			
Trimestru	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Vânzări [mii buc.]	4.6	3.9	5.8	6.3	5.6	5.0	6.6	7.2	5.8	5.4	7.3	7.6	6.1	5.7	7.8	8.2

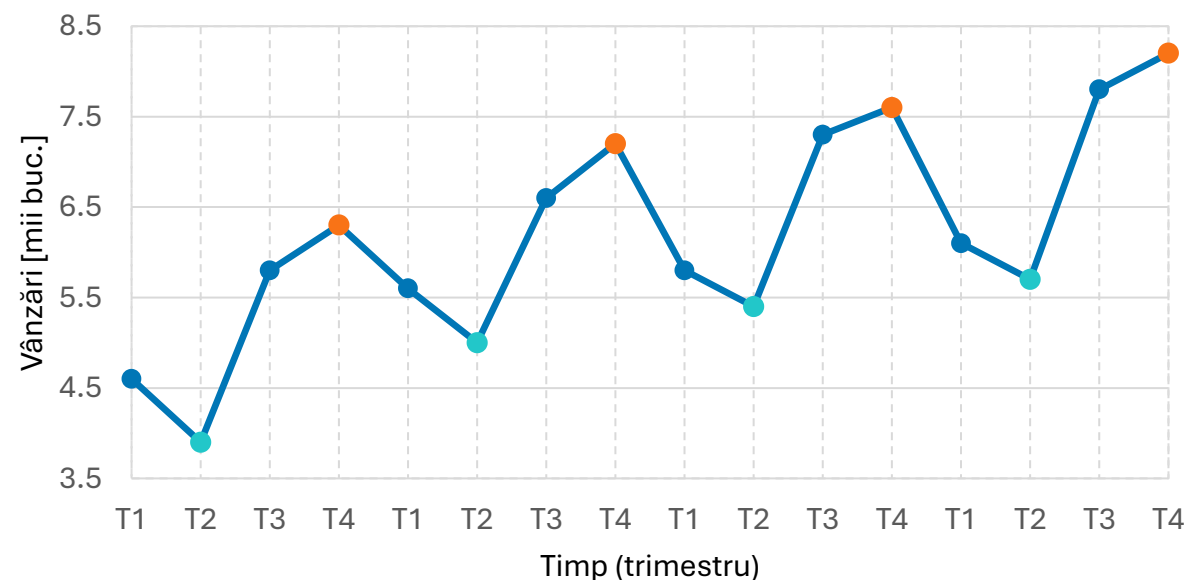
Interpretare

- Seria prezintă un **trend ascendent** de la un an la altul
- Tiparul sezonier se repetă la fiecare **4 trimestre**
- Vânzările sunt, în general, **minime în trimestrul 2**, cresc în trimestrele 3 și 4 și ating nivelul **maxim în trimestrul 4**

Implicație pentru prognoză

Este necesară o metodă de prognoză care să surprindă atât evoluția pe termen lung, cât și variațiile sezoniere

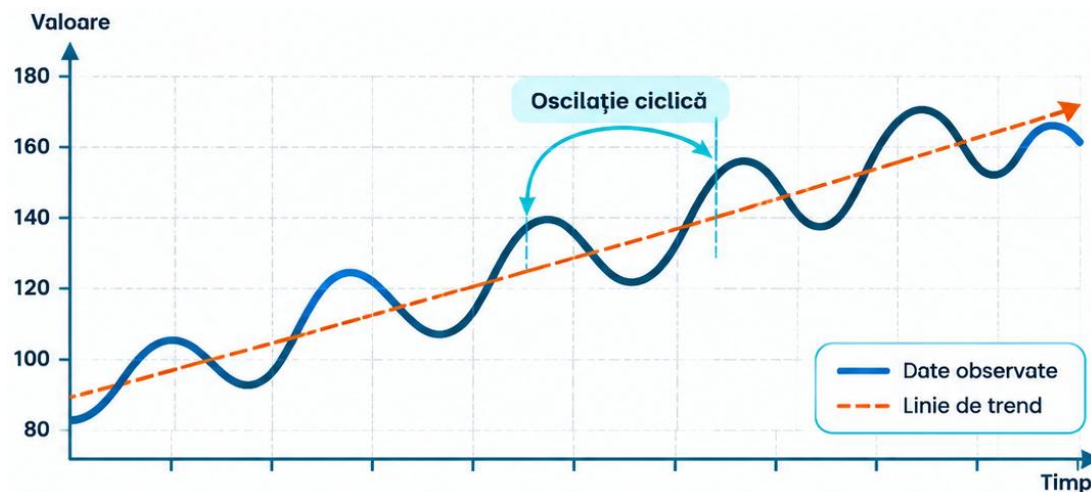
Graficul seriei temporale



Tipar ciclic

Tiparul ciclic (*cyclical pattern*) descrie oscilații de durată variabilă în jurul unei linii de trend, cu perioade succesive deasupra și sub aceasta

- Apare frecvent în seriile economice, fiind asociat cu cicluri de afaceri desfășurate pe mai mulți ani
- Spre deosebire de sezonabilitate, ciclurile nu au o perioadă fixă



Exemplu

Produsul intern brut poate alterna între perioade de **expansiune** și **contractie**, în jurul unui **trend crescător** pe termen lung

Implicație pentru prognoză

Ciclurile de afaceri sunt dificil de prognozat cu precizie

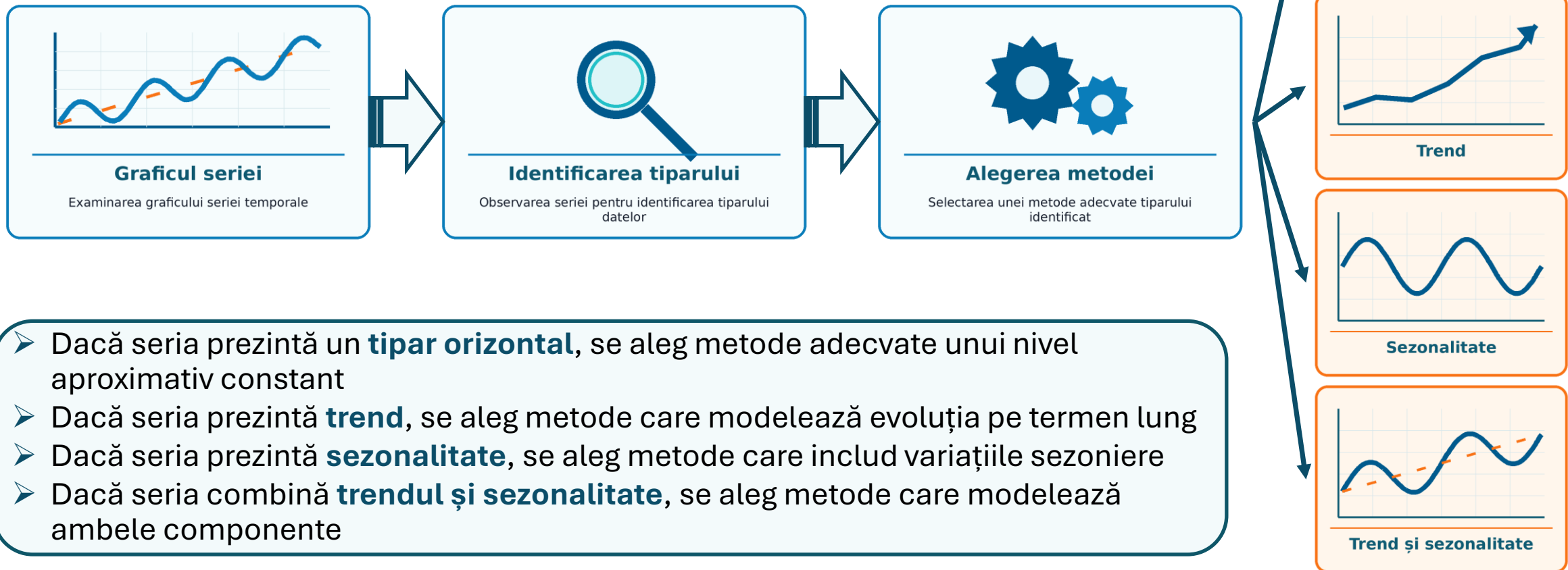
- Componenta ciclică este analizată împreună cu trendul pe termen lung, formând **componenta trend-ciclu**

Ciclic ≠ sezonier

- durată variabilă, de regulă, mai mare de un an
- maximele și minimele nu apar la intervale fixe
- evoluția este dificil de prognozat cu precizie

Selectarea unei metode de prognoză

Primul pas în selectarea metodei de prognoză este examinarea graficului seriei temporale



Cuprins

Analiza previzională

Tipare ale seriilor temporale

Acuratețea prognozei

Metode de prognoză pentru serii cu tipar orizontal

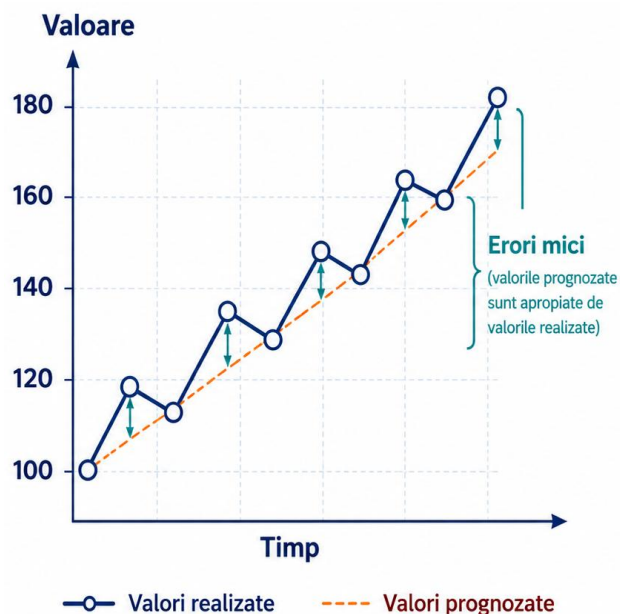
Rezumat

Acuratețea și biasul prognozei

➤ Calitatea unei prognoze poate fi apreciată prin:

Acuratețea prognozei (*forecast accuracy*)

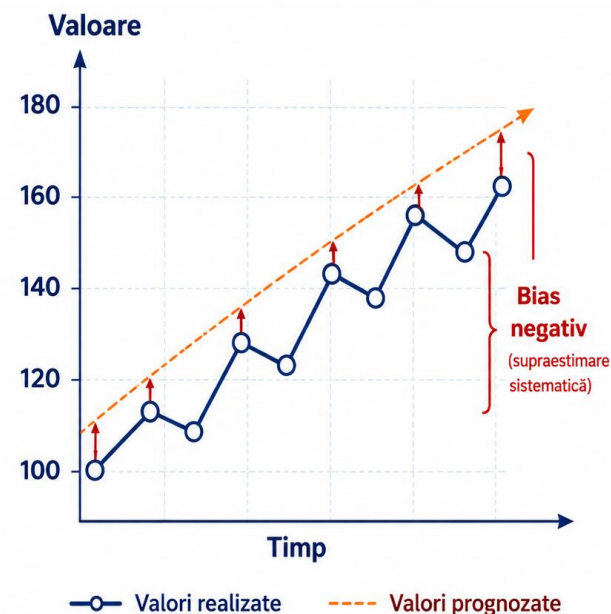
Arată cât de apropiate sunt valorile prognozate de valorile realizate



O prognoză bună are erori mici și un bias apropiat de zero

Biasul prognozei / eroarea sistematică (*forecast bias*)

Arată tendința metodei de a subestima sau supraestima sistematic valorile realizate



Metoda naivă de prognoză

- Revenim la seria temporală a vânzărilor săptămânale de benzină din primul exemplu

Metoda naivă de prognoză

Prognoza pentru perioada t este egală cu valoarea observată în perioada precedentă:

$$F_t = A_{t-1}$$

$$A_1 = 59.5 \Rightarrow F_2 = 59.5$$

$$A_2 = 73.5 \Rightarrow F_3 = 73.5$$

$$A_3 = 66.5 \Rightarrow F_4 = 66.5$$

...

Pentru cele **12 observații** se pot calcula **11 prognoze istorice** — pentru săptămânile 2–12 — și, implicit, **11 erori de prognoză**

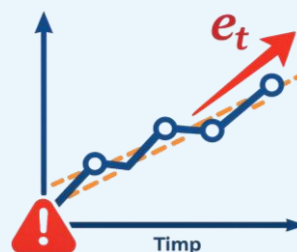
Săptămâna	Vânzări [mii litri]	Prognoză [mii litri]
1	59.5	—
2	73.5	59.5
3	66.5	73.5
4	80.5	66.5
5	63	80.5
6	56	63
7	70	56
8	63	70
9	77	63
10	70	77
11	52.5	70
12	77	52.5

Eroarea de prognoză

Eroarea de prognoză măsoară diferența dintre valoarea realizată și valoarea prognozată

$$e_t = A_t - F_t$$

- A_t – valoarea realizată în perioada t
- F_t – valoarea prognozată în perioada t



Eroare pozitivă – subestimare

$$e_2 = 73.5 - 59.5 = 14 > 0$$

Prognoza a **subestimat** valoarea realizată cu 14 mii de litri

Eroare negativă – supraestimare

$$e_3 = 66.5 - 73.5 = -7 < 0$$

Prognoza a **supraestimat** valoarea realizată cu 7 mii de litri

Mesaj-cheie: interpretarea semnului erorii depinde de convenția adoptată. În acest curs utilizăm:

$$e_t = A_t - F_t$$

Săptămâna	Vânzări [mii litri]	Prognoză [mii litri]
1	59.5	—
2	73.5	59.5
3	66.5	73.5
4	80.5	66.5
5	63	80.5
6	56	63
7	70	56
8	63	70
9	77	63
10	70	77
11	52.5	70
12	77	52.5

Eroarea medie – indicator al biasului

Cum evaluăm performanța prognozelor obținute prin metoda naivă?

Săptămâna	Vânzări [mii litri]	Prognoză [mii litri]	Eroarea de prognoză
1	59.5	—	—
2	73.5	59.5	14
3	66.5	73.5	-7
4	80.5	66.5	14
5	63	80.5	-17.5
6	56	63	-7
7	70	56	14
8	63	70	-7
9	77	63	14
10	70	77	-7
11	52.5	70	-17.5
12	77	52.5	24.5
Total			17.5

- Pentru evaluarea performanței metodei naive calculăm indicatori pe baza celor **11 erori istorice**

Eroarea medie (ME – mean error)

Pentru o serie cu T observații:

$$ME = \frac{1}{T-1} \sum_{t=2}^T (A_t - F_t) = \frac{1}{T-1} \sum_{t=2}^T e_t$$

$$ME = \frac{17.5}{11} = 1.59$$

Interpretare

ME este **pozitivă**, indicând un **bias de subestimare**. Metoda naivă a subestimat vânzările, în medie, cu aproximativ 1.59 mii de litri

Erorile pozitive și negative se pot compensa. Prin urmare, ME măsoară **biasul prognozei**, nu amplitudinea tipică a erorilor

Eroarea medie absolută – indicator al acurateții prognozei

Săptămâna	Vânzări [mii litri]	Prognoză [mii litri]	Eroarea de prognoză	Eroarea absolută
1	59.5	—	—	—
2	73.5	59.5	14	14
3	66.5	73.5	-7	7
4	80.5	66.5	14	14
5	63	80.5	-17.5	17.5
6	56	63	-7	7
7	70	56	14	14
8	63	70	-7	7
9	77	63	14	14
10	70	77	-7	7
11	52.5	70	-17.5	17.5
12	77	52.5	24.5	24.5
Total			17.5	143.5

Eroarea medie absolută (MAE – *mean absolute error*)

$$MAE = \frac{1}{T-1} \sum_{t=2}^T |e_t|$$

$$MAE = \frac{143.5}{11} = 13.05$$

Interpretare

Prognozele metodei naive diferă de valorile realizate cu aproximativ **13.05 mii de litri**, în medie

Avantaje

- evită compensarea erorilor pozitive și negative
- se exprimă în aceeași unitate ca seria
- o valoare MAE mai mică indică o acuratețe mai bună pentru aceeași serie de date

Eroarea medie pătratică – indicator al acurateții prognozei

Săptămâna	Vânzări [mii litri]	Prognoză [mii litri]	Eroarea de prognoză	Eroarea absolută	Eroarea pătratică
1	59.5	—	—	—	—
2	73.5	59.5	14	14	196
3	66.5	73.5	-7	7	49
4	80.5	66.5	14	14	196
5	63	80.5	-17.5	17.5	306.25
6	56	63	-7	7	49
7	70	56	14	14	196
8	63	70	-7	7	49
9	77	63	14	14	196
10	70	77	-7	7	49
11	52.5	70	-17.5	17.5	306.25
12	77	52.5	24.5	24.5	600.25
Total			17.5	143.5	2192.75

MAE și MSE depind de scara datelor și nu sunt potrivite pentru comparații directe între serii exprimate în unități diferite

Eroarea medie pătratică (MSE – mean squared error)

$$MSE = \frac{1}{T-1} \sum_{t=2}^T e_t^2$$

$$MSE = \frac{2192.75}{11} = 199.34$$

Interpretare

Prognozele metodei naive diferă de valorile realizate cu aproximativ **199.34 (mii de litri)²**, în medie

Avantaje

- evită compensarea erorilor pozitive și negative
- penalizează mai puternic erorile mari
- o valoare mai mică indică o prognoză mai bună pentru aceeași serie

Limitare

- se exprimă în unități pătrate și este mai greu de interpretat direct

Rădăcina erorii medii pătratice

Săptămâna	Vânzări [mii litri]	Proгноză [mii litri]	Eroarea de proгноză	Eroarea absolută	Eroarea pătratică
1	59.5	—	—	—	—
2	73.5	59.5	14	14	196
3	66.5	73.5	-7	7	49
4	80.5	66.5	14	14	196
5	63	80.5	-17.5	17.5	306.25
6	56	63	-7	7	49
7	70	56	14	14	196
8	63	70	-7	7	49
9	77	63	14	14	196
10	70	77	-7	7	49
11	52.5	70	-17.5	17.5	306.25
12	77	52.5	24.5	24.5	600.25
Total			17.5	143.5	2192.75

Rădăcina erorii medii pătratice (RMSE – root mean square error)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{T-1} \sum_{t=2}^T e_t^2}$$

$$RMSE = \sqrt{\frac{2192.75}{11}} = \sqrt{199.34} = 14.12$$

Interpretare

RMSE indică o **abatere tipică** de aproximativ **14.12 mii de litri** între valorile prognozate și cele realizate. Deoarece erorile sunt ridicate la pătrat, RMSE penalizează mai puternic erorile mari. Spre deosebire de MSE, RMSE este exprimată în **aceeași unitate de măsură ca seria**, fiind mai ușor de interpretat

Comparație MAE-RMSE: $MAE = 13.05$ mii litri; $RMSE = 14.12$ mii litri

RMSE este mai mare decât MAE deoarece penalizează mai puternic erorile de amplitudine ridicată

Eroarea procentuală absolută medie

Săptămâna	Vânzări [mii litri]	Prognoză [mii litri]	Eroarea de prognoză	Eroarea procentuală [%]	Eroarea procentuală absolută [%]
1	59.5	—	—	—	—
2	73.5	59.5	14	19.05	19.05
3	66.5	73.5	-7	-10.53	10.53
4	80.5	66.5	14	17.39	17.39
5	63	80.5	-17.5	-27.78	27.78
6	56	63	-7	-12.50	12.50
7	70	56	14	20.00	20.00
8	63	70	-7	-11.11	11.11
9	77	63	14	18.18	18.18
10	70	77	-7	-10.00	10.00
11	52.5	70	-17.5	-33.33	33.33
12	77	52.5	24.5	31.82	31.82
Total			17.5	1.19	211.69

Eroarea procentuală absolută (APE – absolute percentage error)

$$APE_t = \left| \frac{A_t - F_t}{A_t} \right| \cdot 100 = \left| \frac{e_t}{A_t} \right| \cdot 100$$

Eroarea procentuală absolută medie (MAPE – mean absolute percentage error)

$$MAPE = \frac{1}{T-1} \sum_{t=2}^T APE_t$$

$$MAPE = \frac{211.69}{11} = 19.24\%$$

Interpretare

Prognozele metodei naive prezintă o eroare procentuală absolută de aproximativ **19.24%**, în medie

Mesaj-cheie

- MAPE este independentă de unitatea de măsură
- permite compararea acurateții metodelor aplicate unor serii cu scări diferite
- nu este definită când valoarea realizată este zero și poate deveni foarte mare pentru valori realizate apropiate de zero

Rezumatul indicatorilor și prognoza viitoare

Folosind **metoda de prognoză naivă** — bazată pe cea mai recentă observație — am obținut următorii **indicatori de evaluare a performanței prognozei**:

$ME = 1.59$ mii litri – bias pozitiv, indicând subestimare

$MAE = 13.05$ mii litri – eroare absolută medie

$MSE = 199.34$ (mii litri)² – penalizează puternic erorile mari

$RMSE = 14.12$ mii litri – abatere tipică, în unitatea seriei

$MAPE = 19.24\%$ – eroare procentuală absolută medie

ME – eroarea medie

MAE – eroarea medie absolută

MSE – eroarea medie pătratică

RMSE – rădăcina erorii medii pătratice

MAPE – eroarea procentuală absolută medie

Prognoza pentru săptămâna 13:

$$F_{13} = A_{12} = 77 \text{ mii litri}$$

- Acuratețea acestei prognoze poate fi evaluată numai după observarea valorii realizate A_{13}

Performanța pe datele istorice oferă o indicație privind comportarea viitoare a metodei, dacă tiparul seriei rămâne relativ stabil. Ea nu garantează însă aceeași acuratețe în viitor

Proгноза prin media cumulativă

- Folosim tot seria temporală a vânzărilor săptămânale de benzină din primul exemplu

Proгноза prin media cumulativă (media tuturor observațiilor anterioare)

Proгноза pentru perioada t este media tuturor valorilor observate până la perioada $t - 1$:

$$F_t = \frac{1}{t-1} \sum_{i=1}^{t-1} A_i, \quad t \geq 2$$

$$F_2 = A_1 = 59.5$$

$$F_3 = \frac{A_1 + A_2}{2} = 66.5$$

$$F_4 = \frac{A_1 + A_2 + A_3}{3} = 66.5$$

...

Săptămâna	Vânzări [mii litri]	Proгноză [mii litri]
1	59.5	—
2	73.5	59.5
3	66.5	73.5
4	80.5	66.5
5	63	80.5
6	56	63
7	70	56
8	63	70
9	77	63
10	70	77
11	52.5	70
12	77	52.5

Evaluarea metodei mediei cumulative

Săptămâna	Vânzări [mii litri]	Prognoză [mii litri]	Eroarea de prognoză	Eroarea absolută	Eroarea pătratică	Eroarea procentuală [%]	Eroarea procentuală absolute [%]
1	59.5	—	—	—	—	—	—
2	73.5	59.50	14.00	14.00	196.00	19.05	19.05
3	66.5	66.50	0.00	0.00	0.00	0.00	0.00
4	80.5	66.50	14.00	14.00	196.00	17.39	17.39
5	63	70.00	−7.00	7.00	49.00	−11.11	11.11
6	56	68.60	−12.60	12.60	158.76	−22.50	22.50
7	70	66.50	3.50	3.50	12.25	5.00	5.00
8	63	67.00	−4.00	4.00	16.00	−6.35	6.35
9	77	66.50	10.50	10.50	110.25	13.64	13.64
10	70	67.67	2.33	2.33	5.44	3.33	3.33
11	52.5	67.90	−15.40	15.40	237.16	−29.33	29.33
12	77	66.50	10.50	10.50	110.25	13.64	13.64
Total			15.83	93.83	1091.11	2.75	141.34

$$ME = \frac{15.83}{11} = 1.44$$

$$MAE = \frac{93.83}{11} = 8.53$$

$$MSE = \frac{1091.11}{11} = 99.19$$

$$RMSE = \sqrt{99.19} = 9.96$$

$$MAPE = \frac{141.34}{11} = 12.85\%$$

Compararea metodei naive cu media cumulativă

Indicator	Metoda naivă	Media cumulativă
ME	1.59	1.44
MAE	13.05	8.53
MSE	199.34	99.19
RMSE	14.12	9.96
MAPE	19.24%	12.85%



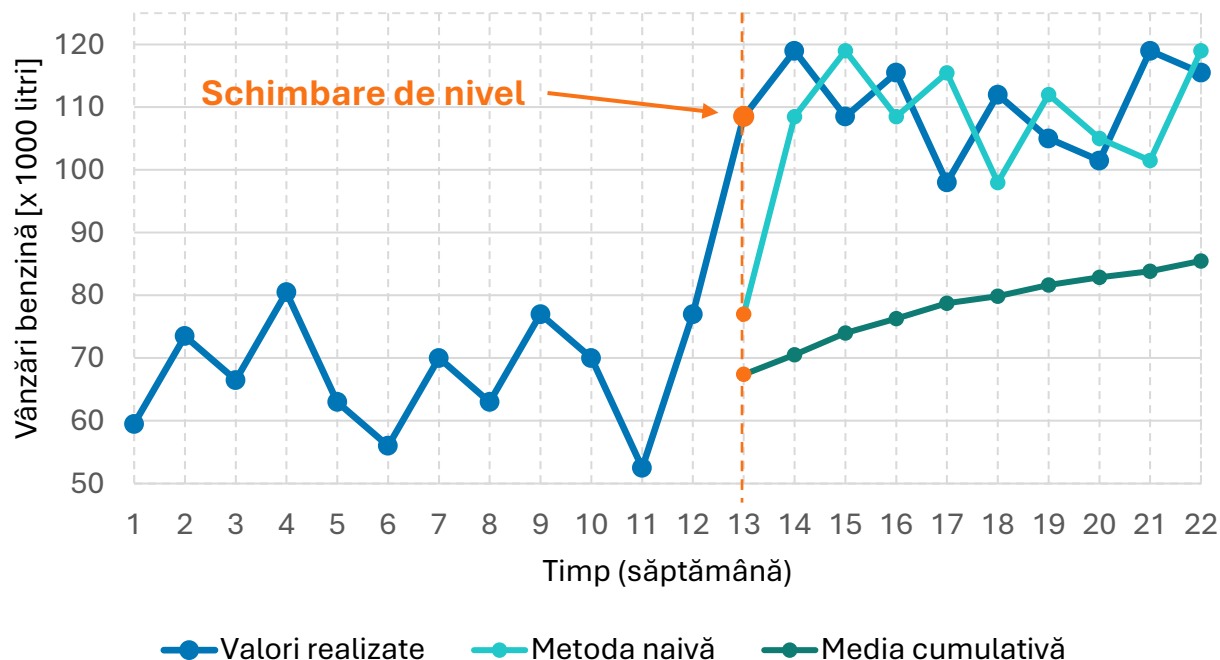
Pentru seria istorică analizată, **media cumulativă are valori mai mici pentru MAE, MSE, RMSE și MAPE**, indicând o acuratețe mai bună decât metoda naivă

Pentru o serie cu **nivel relativ stabil și fără trend sau sezonabilitate**, media cumulativă poate furniza prognoze bune, deoarece atenuează fluctuațiile aleatorii

Adaptarea metodelor de prognoză la o schimbare de nivel

- Revenim la seria temporală a vânzărilor săptămânale de benzină din al doilea exemplu
 - Această serie temporală prezintă o **schimbare structurală de nivel**
- O modificare a condițiilor de afaceri poate produce o **schimbare bruscă a nivelului seriei**, precum cea observată în săptămâna 13

Graficul seriei temporale



Media cumulativă include toate observațiile anterioare schimbării și, prin urmare, se adaptează lent la noul nivel

Metoda naivă utilizează cea mai recentă observație și se adaptează la noul nivel după o singură perioadă. Adaptarea rapidă implică o netezire mai redusă a fluctuațiilor

Acuratețea istorică și alegerea metodei

Indicatorii de acuratețe sunt importanți pentru compararea metodelor de prognoză, dar nu trebuie utilizați izolat

Alegerea metodei de prognoză trebuie să țină cont de:

- tiparul seriei temporale
- orizontul de prognoză
- costul erorilor
- schimbările anticipate ale condițiilor de afaceri
- capacitatea metodei de a se adapta la schimbări

Concluzie

O metodă cu performanțe istorice bune nu garantează aceeași acuratețe dacă tiparul seriei se modifică



În continuare vom analiza metode pentru serii cu tipar orizontal care realizează un compromis între **netezirea fluctuațiilor** și **adaptarea la schimbările de nivel**

Cuprins

Analiza previzională

Tipare ale seriilor temporale

Acuratețea prognozei

Metode de prognoză pentru serii cu tipar orizontal

- Media alunecătoare
- Media alunecătoare ponderată
- Netezirea exponențială

Rezumat

Medii alunecătoare și netezirea exponențială

Trei metode de prognoză adecvate pentru o serie temporală cu un tipar orizontal:

Metoda mediei alunecătoare

Metoda mediei alunecătoare ponderate

Metoda netezirii exponențiale

- Aceste metode realizează un compromis între **netezirea fluctuațiilor** și **adaptarea la schimbările de nivel**

Metoda mediei alunecătoare

Metoda mediei alunecătoare (*moving average method*)

Valoarea prognozată este media aritmetică a celor mai recente k valori realizate

$$F_t = \frac{1}{k} \sum_{j=1}^k A_{t-j} \quad (1)$$

- F_t – prognoza seriei temporale pentru perioada t ($t \geq k + 1$)
- A_{t-j} – observația realizată cu j perioade înainte; $j = 1$ corespunde celei mai recente observații, iar $j = k$ celei mai vechi observații din fereastra de calcul
- k – numărul de valori istorice cu care se calculează media alunecătoare ($k \in \mathbb{N}, k \geq 1$)

Pe măsură ce trece timpul și apar noi valori realizate, ele se introduc în calculul mediei, renunțându-se la valorile cele mai vechi

Alegerea lui k

- k mic → adaptare mai rapidă la schimbări, dar netezire mai redusă
- k mare → netezire mai bună a fluctuațiilor, dar adaptare mai lentă

Alegerea unei valori potrivite pentru k depinde de comportamentul seriei și poate fi validată prin indicatorii de acuratețe calculați pe date istorice

Media alunecătoare cu $k = 3$: exemplu

- Revenim la seria temporală a vânzărilor săptămânale de benzină din primul exemplu

Săptămâna	1	2	3	4	5	6	7	8	9	10	11	12
Vânzări [mii litri]	59.5	73.5	66.5	80.5	63	56	70	63	77	70	52.5	77

Seria prezintă un **tipar orizontal**

Folosim o **medie alunecătoare de ordin $k = 3$**

- Pentru prognoza din săptămâna 4 folosim valorile din săptămânile 1–3

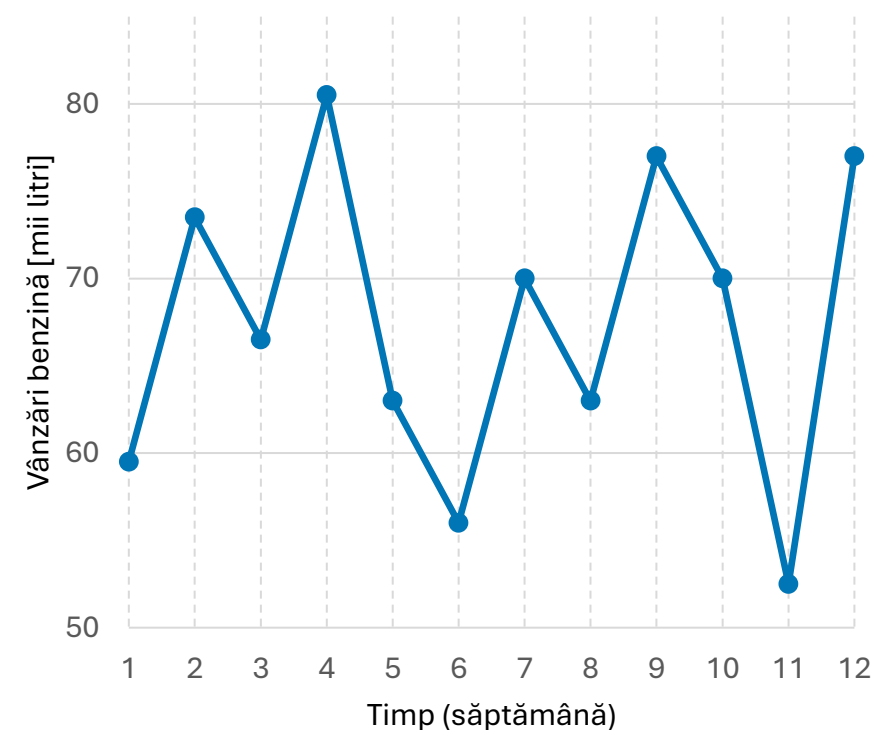
$$F_4 = \frac{59.5 + 73.5 + 66.5}{3} = 66.5$$

- Pentru prognoza din săptămâna 5 folosim valorile din săptămânile 2–4

$$F_5 = \frac{73.5 + 66.5 + 80.5}{3} = 73.5$$

Fereastra de calcul se deplasează cu o perioadă la fiecare nouă prognoză

Graficul seriei temporale



Media alunecătoare cu $k = 3$: exemplu

Săptămâna	Vânzări [mii litri]	Proгноză [mii litri]	Eroarea de proгноză	Eroarea absolută	Eroarea pătratică	Eroarea procentuală [%]	Eroarea procentuală absolută [%]
1	59.5	—	—	—	—	—	—
2	73.5	—	—	—	—	—	—
3	66.5	—	—	—	—	—	—
4	80.5	66.50	14.00	14.00	196.00	17.39	17.39
5	63	73.50	-10.50	10.50	110.25	-16.67	16.67
6	56	70.00	-14.00	14.00	196.00	-25.00	25.00
7	70	66.50	3.50	3.50	12.25	5.00	5.00
8	63	63.00	0.00	0.00	0.00	0.00	0.00
9	77	63.00	14.00	14.00	196.00	18.18	18.18
10	70	70.00	0.00	0.00	0.00	0.00	0.00
11	52.5	70.00	-17.50	17.50	306.25	-33.33	33.33
12	77	66.50	10.50	10.50	110.25	13.64	13.64
Total			0.00	84.00	1127.00	-20.79	129.21

$$ME = \frac{0}{9} = 0.00$$

$$MAE = \frac{84}{9} = 9.33$$

$$MSE = \frac{1127}{9} = 125.22$$

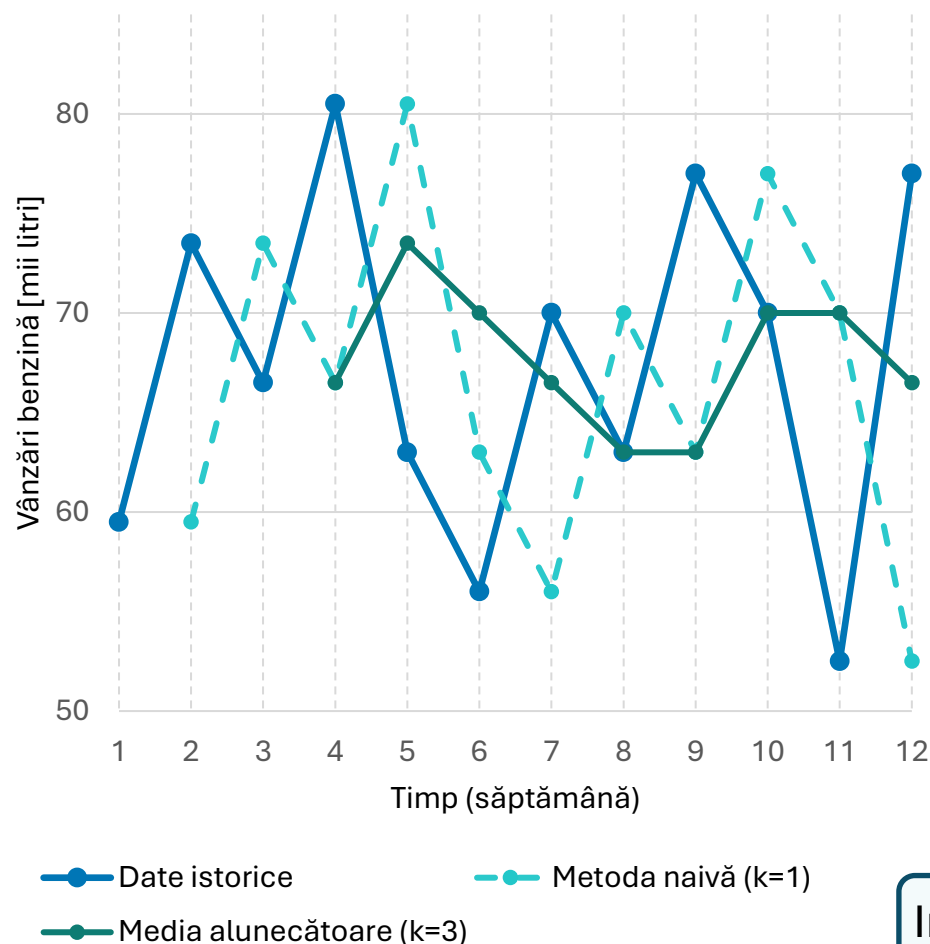
$$RMSE = \sqrt{125.22} = 11.19$$

$$MAPE = \frac{129.21}{9} = 14.36\%$$

ME = 0 indică bias nul pe datele istorice, dar nu înseamnă erori inexistente

Acuratețea prognozei cu medie alunecătoare

Graficul seriei temporale



- Media alunecătoare netezește fluctuațiile aleatorii ale seriei temporale
- Pentru prognoza vânzărilor în săptămâna 13, perioada ce urmează, se calculează media valorilor seriei temporale din săptămânile 10 ÷ 12

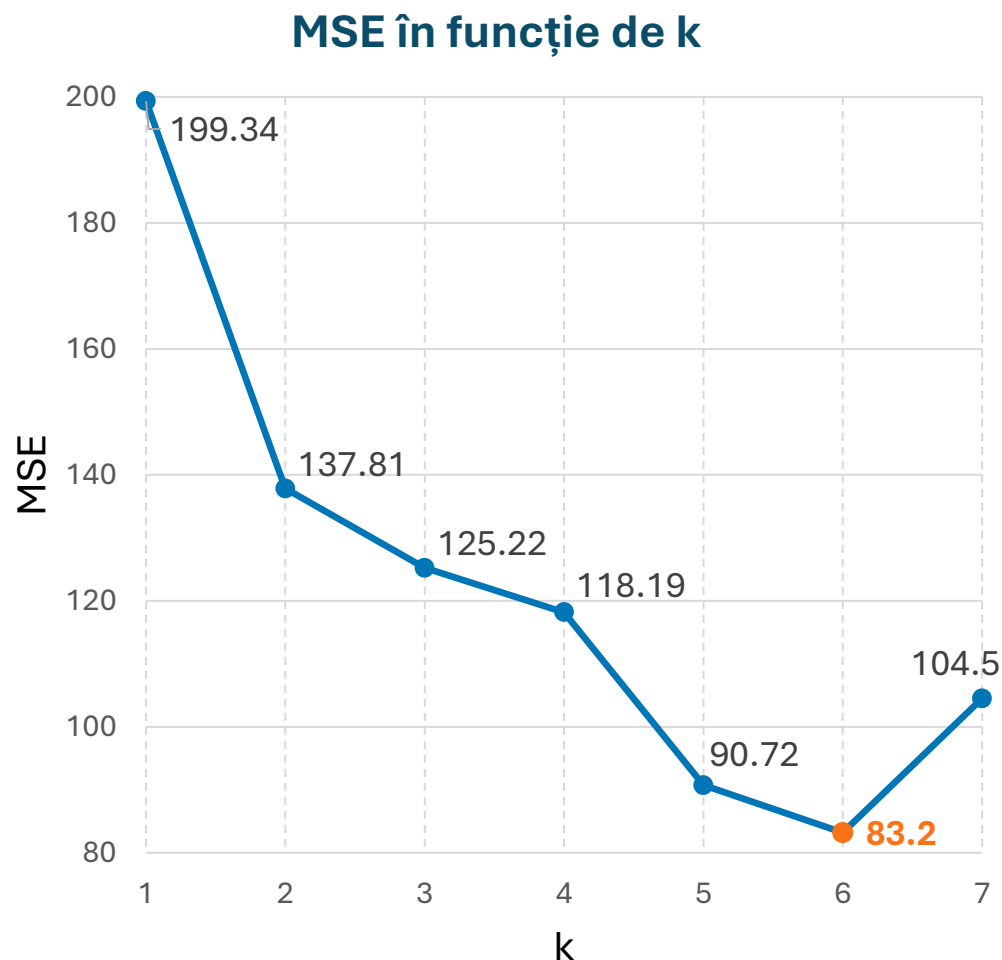
$$F_{13} = \frac{70 + 52.5 + 77}{3} = 66.5$$

Valori mai mici ale indicatorilor de eroare

Indicator	Metoda mediei alunecătoare	
	$k = 1$, săpt. 4–12	$k = 3$, săpt. 4–12
MAE	13.61	9.33
MSE	216.42	125.22
RMSE	14.71	11.19
MAPE	20.23%	14.36%

Indicatorii sunt calculați pentru aceeași perioadă: săptămânile 4–12

Alegerea ordinului mediei alunecătoare



Cum alegem k ?

metoda *trial-and-error*: calculăm MSE pentru mai multe valori ale lui k

Rezultat pentru seria analizată

Pentru seria vânzărilor de benzină, MSE este minim pentru o medie alunecătoare de ordin $k = 6$

Alegerea lui k trebuie validată pe date istorice, dar și interpretată în funcție de comportamentul seriei

$$k = 6, MSE = 83.2$$

Atenție

valoarea optimă pe date istorice nu garantează cea mai bună prognoză viitoare

Când comparăm serii diferite

folosim indicatori relativi, de exemplu **MAPE**

Metoda mediei alunecătoare ponderate

Metoda mediei alunecătoare ponderate (weighted moving average method)

Valoarea prognozată este media ponderată a celor mai recente k valori realizate

$$F_t = \sum_{j=1}^k w_j A_{t-j}, \quad \sum_{j=1}^k w_j = 1 \quad (2)$$

- F_t – prognoza seriei temporale pentru perioada t ($t \geq k + 1$)
- A_{t-j} – observația realizată cu j perioade înainte; $j = 1$ corespunde celei mai recente observații, iar $j = k$ celei mai vechi observații din fereastra de calcul
- w_j – ponderile valorilor istorice care intră în calculul mediei alunecătoare ponderate ($w_j \geq 0$)
- k – numărul de valori istorice cu care se calculează media alunecătoare ponderată ($k \in \mathbb{N}, k \geq 1$)

Pe măsură ce trece timpul și apar noi valori realizate, ele se introduc în calculul mediei, renunțându-se la valorile cele mai vechi

Alegerea ponderilor

- **Ponderi mai mari pentru observațiile recente** → adaptare mai rapidă, dar netezire mai redusă
- **Ponderi mai apropiate ca valoare** → netezire mai bună, dar adaptare mai lentă

Dacă toate ponderile sunt egale, $w_j = 1/k$, metoda devine **media alunecătoare simplă**

Media alunecătoare ponderată cu $k = 3$: exemplu

➤ Revenim la seria temporală a vânzărilor săptămânale de benzină din primul exemplu

Săptămâna	1	2	3	4	5	6	7	8	9	10	11	12
Vânzări [mii litri]	59.5	73.5	66.5	80.5	63	56	70	63	77	70	52.5	77

Pentru prognoză folosim o fereastră de calcul de trei săptămâni ($k = 3$), cu ponderi descrescătoare în funcție de vechimea observațiilor:

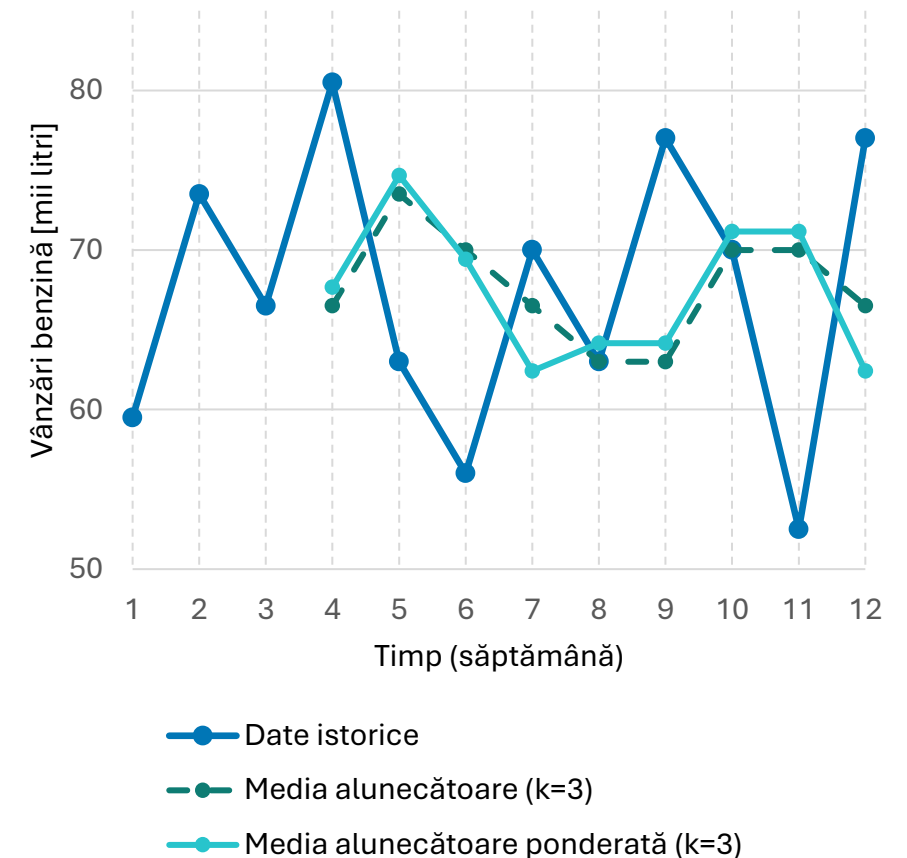
- $w_1 = 3/6$ pentru A_{t-1} , cea mai recentă observație
- $w_2 = 2/6$ pentru A_{t-2}
- $w_3 = 1/6$ pentru A_{t-3} , cea mai veche observație din fereastră

➤ Prognoza din săptămâna 4:

$$F_4 = \frac{3}{6} 66.5 + \frac{2}{6} 73.5 + \frac{1}{6} 59.5 = 67.67$$

Condiție: ponderile descresc odată cu vechimea observațiilor și au suma egală cu 1

Graficul seriei temporale



Evaluarea și alegerea ponderilor

Definirea modelelor candidate

Alegem mai multe combinații ale ordinului k și ale ponderilor w_j , cu:

$$w_j \geq 0, \quad \sum_{j=1}^k w_j = 1$$

Evaluarea performanței

Pentru fiecare combinație calculăm indicatorii de eroare **pe aceeași perioadă de evaluare**: ME, MAE, MSE, RMSE și MAPE

MSE este utilă atunci când dorim să penalizăm mai puternic erorile mari

Atenție la performanța viitoare

Combinația care minimizează eroarea pe datele istorice nu garantează cea mai bună performanță viitoare
Alegerea finală trebuie să realizeze un compromis între **acuratețea istorică, netezirea fluctuațiilor și viteza de adaptare**

Metoda netezirii exponențiale

Netezirea exponențială (*exponential smoothing*)

Reprezintă o familie de metode iterative de prognoză în care observațiilor istorice **li se atribuie ponderi** care descresc exponențial odată cu vechimea lor

- Metodele de netezire exponențială pot modela diferite componente ale unei serii temporale: nivelul, trendul și sezonality

Caz particular: **netezire exponențială simplă** (*simple exponential smoothing*) adecvată seriilor cu tipar orizontal, fără trend și sezonality; utilizează o singură constantă de netezire, α

$$F_{t+1} = \alpha A_t + (1 - \alpha)F_t, \quad 0 \leq \alpha \leq 1 \quad (3)$$

- F_{t+1} – prognoza seriei temporale pentru perioada $t + 1$
- A_t – valoarea realizată pentru perioada t
- F_t – prognoza seriei temporale pentru perioada t
- α – constanta de netezire

Rolul lui α

- α – ponderea celei mai recente valori realizate
- $1 - \alpha$ – ponderea prognozei anterioare
- α controlează compromisul dintre **netezire** și **adaptarea la schimbări**

Netezirea exponențială simplă ca medie ponderată $F_{t+1} = \alpha A_t + (1 - \alpha)F_t$

Ex: Pentru a ilustra metoda netezirii exponențiale considerăm o serie temporală care implică doar trei perioade de date A_1, A_2 și A_3

Pentru inițializare considerăm:

$$F_2 = A_1$$

Proгноza pentru perioada 3

$$F_3 = \alpha A_2 + (1 - \alpha)F_2 = \alpha A_2 + (1 - \alpha)A_1$$

Proгноza pentru perioada 4

$$\begin{aligned} F_4 &= \alpha A_3 + (1 - \alpha)F_3 = \alpha A_3 + (1 - \alpha)[\alpha A_2 + (1 - \alpha)A_1] \\ &= \alpha A_3 + \alpha(1 - \alpha)A_2 + (1 - \alpha)^2 A_1 \end{aligned}$$

- F_4 este o medie ponderată a primelor trei valori ale seriei temporale: A_1, A_2 și A_3
- Suma ponderilor pentru A_1, A_2 și A_3 este egală cu 1

Concluzie

În general, **prognoză F_{t+1}** este o **medie ponderată a tuturor valorilor realizate anterior**, iar **ponderile asociate observațiilor mai vechi scad exponențial**

- **Nu este necesară păstrarea tuturor datelor istorice** pentru actualizarea prognozei
- După alegerea lui α , sunt necesare doar: A_t (cea mai recentă valoare realizată) și F_t (prognoza perioadei curente)

Netezirea exponențială simplă: exemplu

Revenim la seria temporală a vânzărilor săptămânale de benzină din primul exemplu

- **Pentru inițializare** considerăm că prognoza pentru perioada 2 este egală cu valoarea realizată în perioada 1:

$$F_2 = A_1 = 59.5$$

- Valoarea reală a vânzării pentru perioada 2 este 73.5, ca atare eroarea de prognoză pentru perioada 2 este 14 mii litri

Caz particular $\alpha = 0.2$:

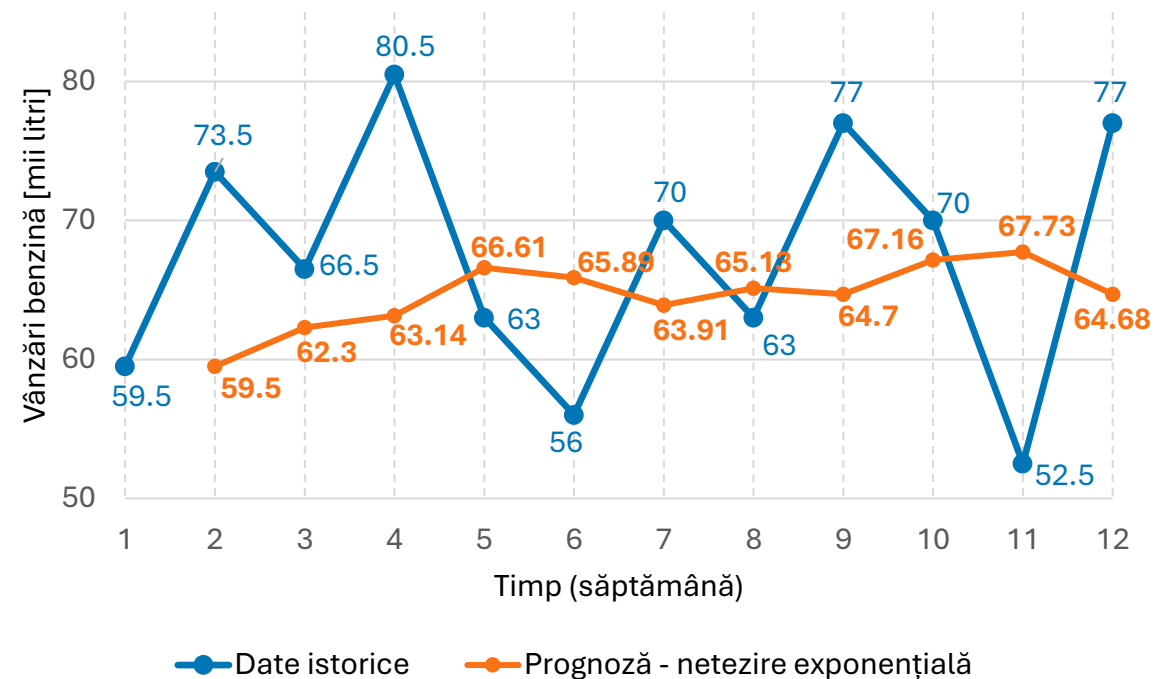
- Prognoza pentru perioada 3 este:
 $F_3 = 0.2A_2 + 0.8F_2 = 0.2 \cdot 73.5 + 0.8 \cdot 59.5 = 62.3$
- Prognoza pentru perioada 4 este:
 $F_4 = 0.2A_3 + 0.8F_3 = 0.2 \cdot 66.5 + 0.8 \cdot 62.3 = 63.14$

Prognoza vânzărilor din săptămâna 13:

$$F_{13} = 0.2A_{12} + 0.8F_{12} = 0.2 \cdot 77 + 0.8 \cdot 64.68 = 67.14$$

- Cu această prognoză, firma își poate face planuri și poate lua anumite decizii

Graficul seriei temporale



Concluzie

Netezirea exponențială reduce fluctuațiile aleatorii ale seriei temporale

Rolul constantei de netezire α

- În evaluările precedente am utilizat constanta de netezire $\alpha = 0.2$
- Deși orice valoare din intervalul $[0, 1]$ este admisibilă, unele valori conduc la prognoze mai bune decât altele
- Forma de corecție a erorii evidențiază rolul constantei de netezire α :

$$F_{t+1} = \alpha A_t + (1 - \alpha)F_t = \alpha A_t + F_t - \alpha F_t$$

$$\Rightarrow F_{t+1} = F_t + \alpha(A_t - F_t)$$

Noua prognoză F_{t+1} este prognoza anterioară F_t , ajustată cu o fracțiune α din **cea mai recentă eroare de prognoză, $A_t - F_t$**

Valoarea lui α poate fi aleasă prin minimizarea unui indicator de eroare calculat pe datele istorice

α mic

- netezire mai puternică
- sensibilitate redusă la fluctuațiile aleatorii
- adaptare mai lentă la schimbările reale de nivel

α mare

- netezire mai redusă
- reacție mai puternică la cea mai recentă observație
- adaptare rapidă la schimbările de nivel, dar sensibilitate mai mare la zgomot

Alegerea constantei de netezire α

Criteriul de alegere

α se alege astfel încât să minimizeze eroarea medie pătratică (MSE) pe datele istorice

Pentru $\alpha = 0.2$

$$MSE = \frac{1210.36}{11} = 110.03$$

Întrebare

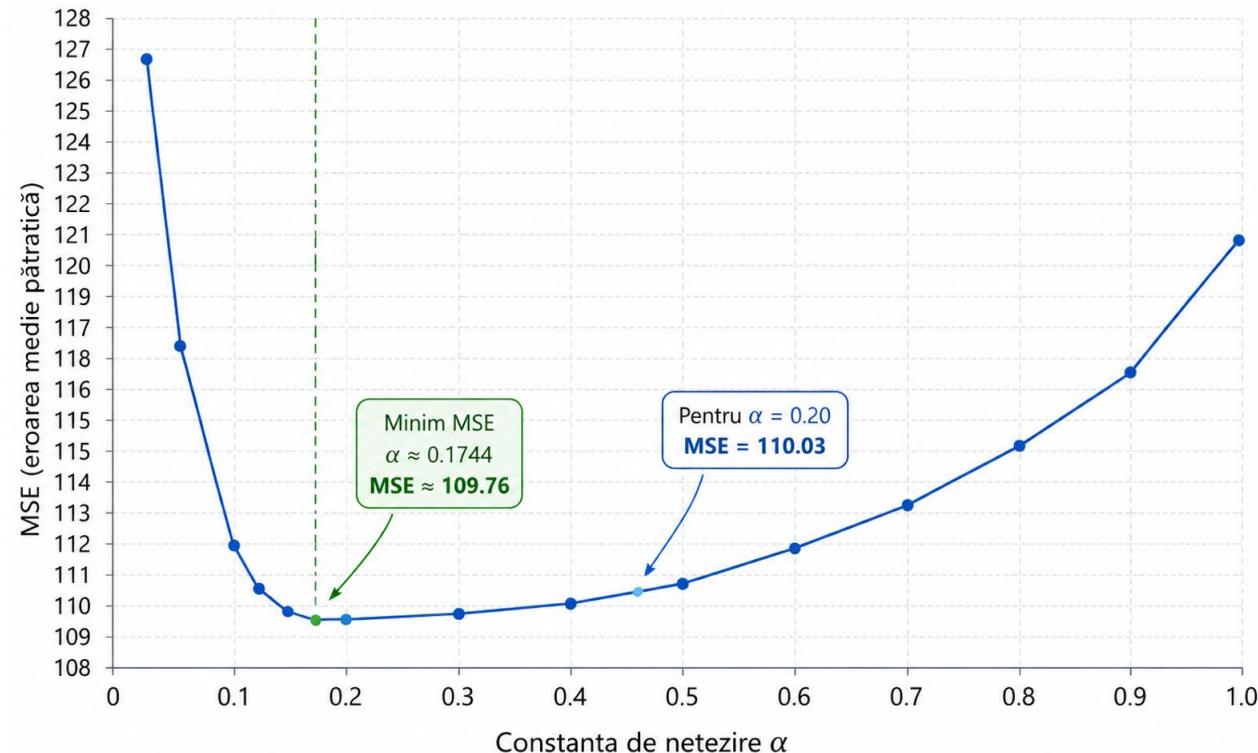
O altă valoare a lui α poate reduce MSE?

Rezultat

MSE este minimă pentru

$$\alpha \approx 0.1744, MSE \approx 109.76$$

Pentru $\alpha = 0.2$, $MSE = 110.03$; diferența este de numai 0.27, adică aproximativ 0.25%



Determinarea lui α cu Excel Solver

A_t – valoarea realizată în perioada t
 F_t – prognoza pentru perioada t

Determinarea valorii lui α care minimizează MSE reprezintă o problemă de **optimizare neliniară cu o singură variabilă de decizie**

Variabila de decizie

α

Funcția obiectiv

$$\min_{\alpha} MSE(\alpha) = \min_{\alpha} \frac{1}{11} \sum_{t=2}^{12} [A_t - F_t]^2$$

Calculul prognozelor

$$F_2 = A_1$$

$$F_t = \alpha A_{t-1} + (1 - \alpha) F_{t-1}; t = \overline{3, 12}$$

Constrângere

$$0 \leq \alpha \leq 1$$

Alpha	0.1744		MSE	109.7586
Săpt	Vânzări [mii litri]	Prognoză [mii litri]	Eroare de prognoză	Eroare pătratic
1	59.5	-	-	-
2	73.5	59.5000	14.0000	196.0000
3	66.5	61.9414	4.5586	20.7805
4	80.5	62.7364	17.7636	315.5457
5	63	65.8342	-2.8342	8.0325
6	56	65.3399	-9.3399	87.2340
7	70	63.7111	6.2889	39.5497
8	63	64.8078	-1.8078	3.2683
9	77	64.4926	12.5074	156.4356
10	70	66.6737	3.3263	11.0641
11	52.5	67.2538	-14.7538	217.6743
12	77	64.6809	12.3191	151.7602
			Total	1207.3448

Soluția obținută

$$\alpha = 0.1744, MSE = 109.76$$

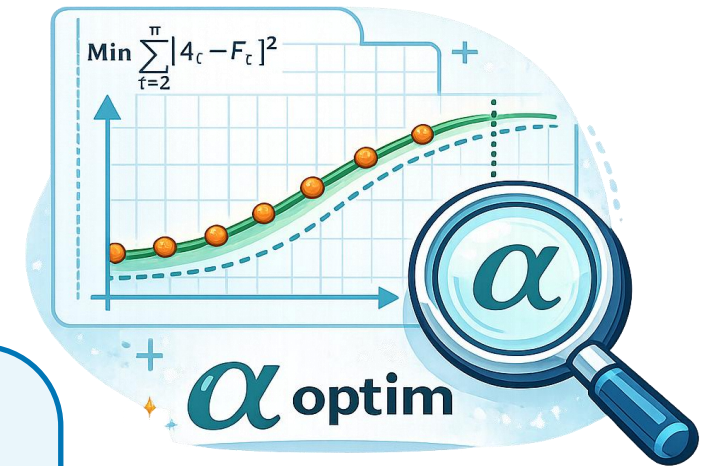
Observație: alegerea inițială $\alpha = 0.2$ este foarte apropiată de valoarea optimă, dar acest lucru nu este garantat în general

Modelul matematic pentru determinarea constantei α

Pentru o serie temporală cu n valori realizate $A_1, A_2, \dots, A_n$, valoarea optimă a constantei de netezire α poate fi obținută prin minimizarea MSE

$$\begin{array}{ll} \text{Min} & \frac{1}{n-1} \sum_{t=2}^n [A_t - F_t]^2 \quad (\text{funcția obiectiv}) \\ \\ \text{s. t.} & F_2 = A_1 \quad (\text{inițializare}) \\ & F_t = \alpha \cdot A_{t-1} + (1 - \alpha)F_{t-1} \quad (\text{actualizarea prognozei}) \\ & 0 \leq \alpha \leq 1 \quad (\text{domeniul constantei de netezire}) \end{array}$$

A_t – valoarea realizată în perioada t
 F_t – prognoza pentru perioada t



Rezumat

Analiza previzională

- ✓ Metode cantitative de prognoză

Tipare ale seriilor temporale

- ✓ Tipar orizontal
- ✓ Tipar de trend
- ✓ Tipar de sezonabilitate / sezonier
- ✓ Tipar de trend și sezonabilitate
- ✓ Tipar ciclic
- ✓ Selectarea unei metode de prognoză

Acuratețea prognozei

Metode de prognoză pentru serii cu tipar orizontal

- ✓ Media alunecătoare
- ✓ Media alunecătoare ponderată
- ✓ Netezirea exponențială

→ Analiza și prognoza seriilor temporale (II)