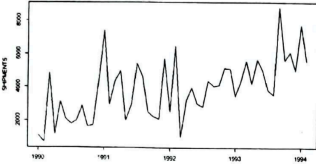


1.1 (0.3) Στο διπλανό διάγραμμα απεικονίζονται οι μηνιαίες αποστολές προϊόντων μιας μεταφορικής εταιρείας. Ποια ποιοτικά χαρακτηριστικά των χρονοσειρών αναγνωρίζετε;



- Μη στασιμότητα: Οι στατιστικές ιδιότητες (μέσος, διακύμανση) δεν παραμένουν σταθερές – παρατηρείται μεταβλητό πλάτος και διαρθρωτικές αλλαγές.
- Εποχιακότητα: Υπάρχουν επαναλαμβανόμενες αιχμές περίπου ανά έτος, που υποδηλώνουν εποχιακά μοτίβα.
- Απουσία κυκλικότητας: Δεν διακρίνονται μακροχρόνιοι κύκλοι πέραν της εποχικότητας.
- Ανοδική τάση: Η σειρά ακολουθεί μακροπρόθεσμα αυξητική πορεία.
- Ασυνέχειες: Εντοπίζονται απότομες αλλαγές (π.χ. αιχμές το 1991-1992), πιθανώς λόγω εξωτερικών γεγονότων.
- Στοχαστικότητα: Παρουσιάζει ακανόνιστες διακυμάνσεις εκτός του εποχιακού μοτίβου, ένδειξη τυχαιότητας ή θορύβου.

1.2 (0.5) «Η ομάδα αναλυτών μιας πολυεθνικής εταιρείας λιανικής πώλησης ηλεκτρονικών ειδών κλήθηκε να προβλέψει τις ετήσιες πωλήσεις ενός νέου τύπου tablet που αναμένεται να κυκλοφορήσει για πρώτη φορά στην ελληνική αγορά. Οι αναλυτές πρότειναν να χρησιμοποιηθεί ως πρόβλεψη **ο μέσος όρος των περασμένων πωλήσεων** tablet της εταιρείας τους στην Ελλάδα, λαμβάνοντας κατά τον υπολογισμό υπόψη ότι τα tablets που εμφανίζουν περιφερή χαρακτηριστικά με το νέο προϊόν. Επίσης, πρότειναν να προσαρμόσουν την πρόβλεψη **βάσει του μεριδίου αγοράς** που σημείωσε το συγκεκριμένο tablet σε άλλες αγορές στις οποίες κυκλοφόρησε νωρίτερα». Ποια μέθοδος παραγωγής κρίσιμων προβλέψεων χρησιμοποιήθηκε; Τι θα συμβουλευάτε εν συνεχεία τους αναλυτές προκειμένου να την εφαρμόσουν αποτελεσματικά;

Πρόβλεψη: **ΜΟ περασμένων πωλήσεων tablet από tablet με περιφερή χαρακτηριστικά**
 ~ **ΜΟ: στατική μέθοδος**

Προσαρμογή: **βάσει μεριδίου αγοράς του tablet σε άλλες αγορές νωρίτερα**
 ~ **Αναλογία (σελ. 254)**, δηλαδή διαδικασία κρίσιμης πρόβλεψης

Προτείνεται συνδυασμός των μεθόδων

1.3 (0.2) Δίνεται η παρακάτω τριμηνιαία εποχιακή χρονοσειρά. Προβείτε στις κατάλληλες ενέργειες για να συμπληρώσετε τις τιμές που λείπουν (NA) αιτιολογώντας σύντομα τις επιλογές σας.

Τρίμηνο	Q1-19	Q2-19	Q3-19	Q4-19	Q1-20	Q2-20	Q3-20	Q4-20	Q1-21	Q2-21
Τιμή	802	NA	658	924	NA	622	674	1052	858	635

← έχει $Q_1 \rightarrow$ ίσως $\exists$ εποχικότητα

	Q1	Q2	Q3	Q4
19	802	NA	658	924
20	NA	622	674	1052
21	858	635		

→ **έκκαθαρη εποχικότητα:**

$$Q_2-19 = \text{Avg}(622, 635) = 628.5$$

$$Q_1-20 = \text{Avg}(802, 858) = 830$$

1.4 (0.4) Η ακόλουθη χρονοσειρά παρουσιάζει τις μηνιαίες πωλήσεις ενός νέου προϊόντος καθαρισμού. Η εταιρεία που το παράγει το προώθησε ενεργά τον τρίτο μήνα της κυκλοφορίας του μέσω τηλεοπτικής διαφήμισης. Το ίδιο σκοπεύει να κάνει και τον επόμενο μήνα. Υπολογίστε τον αντίκτυπο της προωθητικής ενέργειας και προβλέψτε τις πωλήσεις του 8ου μήνα χρησιμοποιώντας ως μοντέλο έναν ΚΜΟ(3). (σελ. 9)

προώθηση

Μήνας	M1	M2	M3	M4	M5	M6	M7
Τιμή	525	536	704	555	498	527	496

$$D_3 = 704$$

$$D_3' = D_2 = 536 \text{ (οχι zion)}$$

$$\text{Impact} = \frac{D_3 - D_3'}{D_3'} \cdot 100\% = 31.34\%$$

$$F_8 = \text{Avg}(F_5, F_6, F_7) = 507 \text{ (χωρίς προώθηση)}$$

$$F_8' = F_8 \cdot (1 + \text{Impact}) = 665.91 \approx 666 \text{ (με προώθ)}$$

1.5 (0.4) Αντιστοιχίστε τα διαγράμματα 1, 2, 3 και 4 με τις τιμές του συντελεστή γραμμικής συσχέτισης στα δεξιά (a-f):

Διάγραμμα 1



Διάγραμμα 2



Διάγραμμα 3



Διάγραμμα 4



a. $r = 0.9$
b. $r = -0.9$
c. $r = 0.75$

d. $r = 0.5$
e. $r = 0.25$
f. $r = -0.5$

σελ 40

Γ: γραμμική συσχέτιση

$0.8 < |r| < 1$: πολύ ισχυρή Γ

$0.7 < |r| < 0.8$: ισχυρή Γ

$0.5 < |r| < 0.7$: μέση Γ

$0.3 < |r| < 0.5$: ασθενής Γ

$0 < |r| < 0.3$: δεν υπάρχει Γ

$r > 0$: αυτoυσα
 $r < 0$: φθίνουσα

Διαγ 1: πολύ ισχυρή Γ + 1 $\Rightarrow b_r = -0.9$

Διαγ 2: ισχυρή Γ + 1 $\Rightarrow c_r = 0.75$

Διαγ 3: μέση Γ + 1 $\Rightarrow d_r = 0.5$

Διαγ 4: δεν υπάρχει Γ $\Rightarrow e_r = 0.25$

1.6 (0.2) Αν οι προβλέψεις ενός μοντέλου είναι $F = \{10, 9, 7\}$ και οι πραγματικές τιμές της χρονοσειράς είναι $Y = \{8, 7, 8\}$, αποφασίστε σχετικά με την προκατάληψη των προβλέψεων.

t	y_t	F_t	e_t
1	8	10	-2
2	7	9	-2
3	8	7	1

προκατάληψη $\Rightarrow$ ΜΕ

$ME = \frac{1}{n} \sum (y_t - F_t) = \frac{1}{n} \sum e_t = -1 < 0$ απιστοδορία

ΘΕΜΑ 2ο (2.5 μονάδες)

Δίνεται η παρακάτω ετήσια χρονοσειρά, οι τιμές της οποίας παρουσιάζουν τα έσοδα ενός οργανισμού:

Έτος	2013	2014	2015	2016	2017	2018	2019	2020
Τιμή (εκ. €)	18	19	22	20	24	27	25	29

$\rightarrow$ γραμμική ανοδική τάση

2.1 (1.0) Δώστε τις προβλέψεις σας για τα έτη 2021 και 2022 χρησιμοποιώντας τη μέθοδο της Εκθετικής Εξομάλυνσης Γραμμικής Τάσης $\alpha = 0.1 + AM/5$, $\beta = \alpha - 0.05$ και αρχικά επίπεδα της επιλογής σας, όπου AM το τελευταίο ψηφίο του αριθμού μητρώου σας.

Holt Exponential Smoothing (σελ. 143)

► αρχικό επίπεδο S_0 : ΜΟ όλων / n πρώτων παρατηρήσεων

επιλέγω αυθαίρετα $n = 3 \Rightarrow S_0 = \frac{\sum y_t}{3} = 19.67$

► αρχική τάση T_0 : διαφορά $y_1 - y_2$ ή $(y_n - y_1) / (n - 1)$

πάλι επιλέγω αυθαίρετα $n = 8: T_0 = \frac{y_8 - y_1}{7} = 1.57$

Holt: $e_t = y_t - F_t$

$\alpha = 0.12$

$S_t = S_{t-1} + T_{t-1} + \alpha \cdot e_t$

$\beta = 0.07$

$T_t = T_{t-1} + \alpha \cdot \beta \cdot e_t$

$F_{t+m} = S_t + m \cdot T_t$

t	Y _t	F _t	e _t	S ₀	T ₀
0				19.67	1.57
1	18	21.24	-3.24	20.85	1.54
2	19	22.39	-3.39	21.99	1.52
3	22	23.50	-1.50	23.32	1.50
4	20	24.83	-4.83	24.25	1.46
5	24	25.71	-1.71	25.50	1.45
6	27	26.95	0.05	26.96	1.45
7	25	28.41	-3.41	28.00	1.42
8	29	29.42	-0.42	29.37	1.42

Holt: $e_t = Y_t - F_t$ $a = 0.12$
 $S_t = S_{t-1} + T_{t-1} + a \cdot e_t$ $b = 0.07$
 $T_t = T_{t-1} + a \cdot b \cdot e_t$ $S_0 = 19.67$
 $F_{t+m} = S_t + m \cdot T_t$ $T_0 = 1.57$

t=9: $F_9 = S_8 + 1 \cdot T_8 = 29.37 + 1 \cdot 1.42 = 30.78$

t=10: $F_{10} = S_8 + 2 \cdot T_8 = 29.37 + 2 \cdot 1.42 = 32.20$

2.2 (0.2) Ποιες είναι οι προβλέψεις της αφελούς μεθόδου για τα ίδια έτη;

Naive: $F_9 = F_0 = Y_8 = 29$

2.3 (0.3) Δεδομένου ότι οι πραγματικές τιμές στα αντίστοιχα έτη πρόβλεψης ήταν 32 και 34 εκ. Ευρώ, αξιολογήστε την ακρίβεια των δύο μεθόδων. Πώς αιτιολογείται ποιοτικά η διαφορά τους;

t	Y _t	Holt F _t	Holt e _t	Holt e _t ²	t	Y _t	Naive F _t	Naive e _t	Naive e _t ²
8	32	30.78	1.22	1.49	8	32	29	3	9
9	34	32.20	1.70	2.89	9	34	29	5	25

$MSE_{naive} = \frac{1}{n} \sum e_t^2 = 17 \rightarrow$ δεν υπολ τίσση

$MSE_{holt} = \frac{1}{n} \sum e_t^2 = 2.19$

ιδανική για βραχυπρόθ πρόβλεψη κσ
 γραμμικής σχεδόν τίσσης

2.4 (1.0) Αν οι συντελεστές του μοντέλου απλής γραμμικής παλινδρόμησης είναι $a = 16.25$ και $b = 1.5$, αποφανθείτε αν αυτοί είναι στατιστικά σημαντικοί για επίπεδο εμπιστοσύνης 90%. (σελ 182)

URL: Εξίσωση: $F_t = a + b \cdot t$ $\bar{t} = \text{Avg}(1, 2, 3, 4, 5, 6, 7, 8) = 4.5$

τυπικό σφάλμα συντελεστών:

$SE_{a-} = \hat{\sigma}_e \cdot \sqrt{\frac{1}{n} + \frac{\bar{t}^2}{\sum (t_i - \bar{t})^2}}$ $SE_{b-} = \hat{\sigma}_e \cdot \sqrt{\frac{1}{\sum (t_i - \bar{t})^2}}$

τυπική απόκλιση:

$\hat{\sigma}_e = \sqrt{\frac{\sum (Y_i - F_i)^2}{n-k}}$

t	Y _t	F _t	e _t	e _t ²	t - $\bar{t}$	(t - $\bar{t}$) ²
1	18	17.75	0.25	0.06	-3.5	12.25
2	19	19.25	-0.25	0.06	-2.5	6.25
3	22	20.75	1.25	1.56	-1.5	2.25
4	20	22.25	-2.25	5.06	-0.5	0.25
5	24	23.75	0.25	0.06	0.5	0.25
6	27	25.25	1.75	3.06	1.5	2.25
7	25	26.75	-1.75	3.06	2.5	6.25
8	29	28.25	0.75	0.56	3.5	12.25
				13.5		42.00

df = n - k = 8 - 2 = 6

conf level: p = 90%

⇒ πρέπει $t_{obs} > t_{90}$

για να είναι

στατιστικά σημαντ

(n-k)	df	0.1	0.05	0.025	0.01	0.005
1	3.08	6.31	12.71	31.82	63.66	
2	1.89	2.92	4.30	6.96	9.92	
3	1.64	2.35	3.18	4.54	5.84	
4	1.53	2.13	2.78	3.75	4.60	
5	1.48	2.02	2.57	3.36	4.03	
6	1.44	1.94	2.45	3.14	3.71	
7	1.41	1.89	2.36	3.00	3.50	
8	1.40	1.86	2.31	2.90	3.36	
9	1.38	1.83	2.26	2.82	3.25	
10	1.37	1.81	2.23	2.76	3.17	
11	1.36	1.80	2.20	2.72	3.11	
12	1.36	1.78	2.18	2.68	3.05	
13	1.35	1.77	2.16	2.65	3.01	
14	1.34	1.76	2.14	2.62	2.98	
15	1.34	1.75	2.13	2.60	2.95	
16	1.34	1.75	2.12	2.58	2.92	
17	1.33	1.74	2.11	2.57	2.90	
18	1.33	1.73	2.10	2.55	2.88	
19	1.33	1.73	2.09	2.54	2.86	
20	1.33	1.72	2.09	2.53	2.85	
21	1.32	1.72	2.08	2.52	2.83	
22	1.32	1.72	2.07	2.51	2.82	
23	1.32	1.71	2.07	2.50	2.81	
24	1.32	1.71	2.06	2.49	2.80	
25	1.32	1.71	2.06	2.49	2.79	
30	1.31	1.70	2.04	2.46	2.75	
40	1.30	1.68	2.02	2.42	2.70	
50	1.30	1.68	2.01	2.40	2.68	
60	1.30	1.67	2.00	2.39	2.66	
70	1.29	1.67	1.99	2.38	2.65	
80	1.29	1.66	1.99	2.37	2.64	
90	1.29	1.66	1.99	2.37	2.63	
100	1.29	1.66	1.98	2.36	2.63	
∞	1.28	1.64	1.96	2.33	2.58	

Confidence level

$$\hat{\sigma}_e = \sqrt{\frac{135}{8-2}} = 1.5 \quad (k=2 \text{ παλιν}). \quad \hat{\sigma}_{\epsilon a} = 1.5 \sqrt{\frac{1}{8} + \frac{4.5^2}{42}} = 1.17, \quad \hat{\sigma}_{\epsilon b} = 1.5 \sqrt{\frac{1}{42}} = 0.23$$

στατιστικοί δείκτες: $a'=b'=0$

$$t_a = \frac{a-a'}{\hat{\sigma}_{\epsilon a}}, \quad t_b = \frac{b-b'}{\hat{\sigma}_{\epsilon b}} \Rightarrow t_a = \frac{1625}{1.17} = 13.9, \quad t_b = \frac{15}{0.23} = 6.48 \rightarrow t_a, t_b > 1.94 \rightarrow \text{απαρ σημαντ}$$

ΘΕΜΑ 3ο (2.0 μονάδες)

Δίνεται χρονοσειρά που παρουσιάζει τη τετραμηνιαία ζήτηση ανταλλακτικών μιας γεωργικής μηχανής:

Τετράμηνο	P1-19	P2-19	P3-19	P1-20	P2-20	P3-20	P1-21	P2-21	P3-21	P1-22	P2-22
Τιμή	0	38	0	0	25	0	0	32	0	0	29

→ 3 περί ΕΠΟΧΙΚΟΤΗΤΑ

3.1 (1.0) Δώστε τις προβλέψεις σας για το επόμενο έτος χρησιμοποιώντας τη μέθοδο του Croston.

Croston: (σελ 227)

$$S_0 = \bar{y}_t$$

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

$$a = 0.05$$

$$e_t = y_t - F_t$$

$$F_{\text{Croston}} = \frac{F_{\text{dem}}}{F_{\text{int}}}$$

$$S_t = S_{t-1} + a \cdot e_t$$

$$S_0^{\text{dem}} = \bar{y}_{\text{dem}} = 31$$

$$S_0^{\text{int}} = \bar{y}_{\text{int}} = 2.75$$

$$F_{\text{Croston}} = \frac{30.98}{2.75} = 11.25 = F_{11} = F_{12} = F_{13}$$

Demands

y_t	F_t	e_t	S_t
-	-	-	31
38	31	7	31.35
25	31.35	-6.35	31.03
32	31.03	0.97	31.08
29	31.08	-2.08	30.98
	30.98		

Intervals

y_t	F_t	e_t	S_t
-	-	-	2.75
2	2.75	-0.75	2.71
3	2.71	0.29	2.73
3	2.73	0.27	2.74
3	2.74	0.26	2.75
	2.75		

3.2 (1.0) Συμφωνείτε με τη χρήση της παραπάνω μεθόδου βάσει των ποιοτικών χαρακτηριστικών που παρουσιάζει η εν λόγω χρονοσειρά; Προτείνετε μία προσέγγιση που θεωρείτε καταλληλότερη.

	P1	P2	P3
19	0	38	0
20	0	25	0
21	0	32	0
22	0	29	0

Γεμάταρη εποχικότητα: ⇒ Croston άσχετη

μέσος όρος αντιστοιχων περιόδων:

$$P3-22: F_n = 0$$

$$P1-23: F_{12} = 0$$

$$P2-23: F_{13} = \text{Avg}(38, 25, 32, 29) = 31$$

ΘΕΜΑ 4ο (3.5 μονάδες)

Δίνεται η ακόλουθη εποχιακή τριμηνιαία χρονοσειρά:

Τρίμηνο	Q1-19	Q2-19	Q3-19	Q4-19	Q1-20	Q2-20	Q3-20
Τιμή	208	363	343	253	218	376	272

4.1 (2.0) Εφαρμόστε τη μέθοδο πρόβλεψης Theta για την πρόβλεψη των επόμενων δύο περιόδων.

Χρησιμοποιείτε δύο γραμμές Theta με παραμέτρους $\theta_1=0$ και $\theta_2=2.5$. Για την προέκταση της γραμμής με παράμετρο θ_2 , χρησιμοποιείτε ως μέθοδο πρόβλεψης έναν ΚΜΟ(2).

Δίνονται οι δείκτες εποχικότητας των πρώτων τριών τριμήνων ($SI_1=0.71$, $SI_2=1.24$, $SI_3=1.17$) και η εξίσωση της ευθείας γραμμικής παλινδρόμησης που περνά βέλτιστα από τα αποεποχικοποιημένα δεδομένα: $Y=308-5.2t$

$$2SI_3 = 4 \Rightarrow 0.71 + 1.24 + 1.17 + x = 4 \Rightarrow SI_4 = 0.88$$

Βήμα 1: ανοσοποίησηση $Y_t' = Y_t / S_t$

Βήμα 2: ανοσοποίησηση $Y_t^{0:0} = 308 - 5.2t$, $Y_t^{0:25} = 2.5 Y_t - 1.5 Y_t^{0:0}$

Βήμα 3: Πρόβλεψη: $\theta = 0$: απλή γραμμή παλινδρ, $\theta = 2.5$: KMO (2)

Βήμα 4: Συνδυασμός: $F_t' = (Y_t^{0:0} + Y_t^{0:25}) / 2$

Βήμα 5: Ενομοποίησηση: $F_t = F_t' \cdot S_t$

			①	②	③	④	⑤	σ ^α 4.2	
t	Y _t	S _t	Y _t '	Y _t ^{0:0}	Y _t ^{0:25}	F _t '	F _t	e _t	e _t ²
1	208	0.71	292.96	302.8	278.19	290.5	206.25	175	3.05
2	363	1.24	292.74	297.6	285.45	291.53	361.49	151	2.27
3	343	1.17	293.16	292.4	294.31	293.35	343.22	-0.22	0.05
4	253	0.88	289.50	283.2	289.95	287.58	253.07	-0.07	0
5	218	0.71	307.04	282.0	344.61	313.30	222.45	-4.45	19.76
6	376	1.24	303.23	276.8	342.86	309.83	384.19	-8.19	67.11
7	272	1.17	232.48	271.6	173.80	222.70	260.56	11.44	130.94
8		0.88		266.4	258.33	262.37	230.88		
9		0.71		261.2	258.33	259.77	184.43		

Τελικές προβλέψεις:

$$F_9 = 230.88$$

$$F_{10} = 184.43$$

$$RMSE = \sqrt{\frac{1}{n} \sum e_t^2} = \sqrt{31.88} = 5.647$$

$$AVG = 31.88$$

$$\rightarrow t = 1.645$$

4.2 (1.0) Δώστε το διάστημα εμπιστοσύνης των προβλέψεων σας θεωρώντας επίπεδο 90%.

$$F_t = F_t + t \cdot RMSE \cdot \sqrt{1-n} \quad \left\{ \begin{array}{l} F_9 = 230.88 \pm 1.645 \cdot 5.647 \cdot \sqrt{1} = 230.88 \pm 9.29 \\ F_{10} = 184.43 \pm 1.645 \cdot 5.647 \cdot \sqrt{2} = 184.43 \pm 13.14 \end{array} \right.$$

4.3 (0.5) Τρία μοντέλα ARIMA χρησιμοποιούνται για την πρόβλεψη της ίδιας χρονοσειράς. Το πρώτο μοντέλο, ARIMA(1,1,0), εμφανίζει προσδοκώμενη πιθανότητα $\log L = -50$. Τα άλλα δύο μοντέλα, ARIMA(0,1,1) χωρίς σταθερά και ARIMA(0,1,1) με σταθερά, εμφανίζουν πιθανότητα -52 και -49.5 αντίστοιχα. Ποιο μοντέλο θα επιλέγατε για την παραγωγή προβλέψεων;

$$AIC = -2 \log L + 2(p+q+p+q+k+1) = -2 \log L + 2(p+q+k+1) \quad \downarrow AIC \rightarrow \text{καλύτερο μοντέλο}$$

► χωρίς σταθερά ($k=0$)

$$ARIMA(1,1,0) \text{ χωρίς } c: \log L = -50 \rightarrow AIC = -2(-50) + 2(1+0+0+1) = 100+4 = 104 \sim \text{καλύτερο}$$

$$ARIMA(0,1,1) \text{ χωρίς } c: \log L = -52 \rightarrow AIC = -2(-52) + 2(1+0+0+1) = 104+4 = 108$$

► με σταθερά ($k=1$)

$$ARIMA(1,1,0) \text{ με } c: \log L = -50 \rightarrow AIC = -2(-50) + 2(1+0+1+1) = 100+6 = 106$$

$$ARIMA(0,1,1) \text{ με } c: \log L = -49 \rightarrow AIC = -2(-49) + 2(1+0+1+1) = 98+6 = 105 \sim \text{καλύτερο}$$