

Statisztika 2

példatár

mateking.hu

6. IDŐSOROK ELEMZÉSE

6.1. Egy részvény árfolyamának alakulását 20 napig figyeltük. Illesszünk az adatokra három napos mozgóátlagos trendet, majd lineáris trendet. Számítsuk ki a változás átlagos napi mértékét és hasonlítsuk össze a lineáris trend megfelelő paraméterével.

nap	Részvény ára (USD)
1.	90
2.	91
3.	88
4.	87
5.	87
6.	86
7.	88
8.	91
9.	93
10.	94
11.	93
12.	95
13.	97
14.	98
15.	97
16.	100
17.	99
18.	102
19.	98
20.	95

6.2. Egy új termék piacra történő bevezetésének adatai az alábbiak voltak.

év	1000 emberből a termékkel rendelkezők száma			
	I. negyedév	II. negyedév	III. negyedév	IV. negyedév
2008	$y_1 = 10$	$y_2 = 12$	$y_3 = 14$	$y_4 = 15$
2009	$y_5 = 17$	$y_6 = 19$	$y_7 = 20$	$y_8 = 21$
2010	$y_9 = 23$	$y_{10} = 25$	$y_{11} = 28$	$y_{12} = 30$
2011	$y_{13} = 35$	$y_{14} = 39$	$y_{15} = 43$	$y_{16} = 46$

Illesszünk az adatokra lineáris, majd exponenciális trendet és döntsük el, hogy melyik illeszkedik jobban. Mindkét esetben vizsgáljuk meg a szezonalitást.

6.3. Egy üzem termelése három egymást követő évben az alábbiak szerint alakult. Illesszünk az adatsorra lineáris majd exponenciális trendet, vizsgáljuk meg, hogy melyik illeszkedik jobban, és adjuk meg a szezonalitást.

ÉV		termelés (1000 tonna)
2011	TÉL	120
	TAVASZ	142
	NYÁR	164
	ŐSZ	196
2012	TÉL	240
	TAVASZ	256
	NYÁR	324
	ŐSZ	360
2013	TÉL	420
	TAVASZ	512
	NYÁR	576
	ŐSZ	600

6.4. Egy üzem sörtermelése három egymást követő évben az alábbiak szerint alakult. Illesszünk az adatsorra lineáris majd exponenciális trendet, vizsgáljuk meg, hogy melyik illeszkedik jobban, és adjuk becslést a következő év sörtermelésére a jobban illeszkedő trend alapján, szezonalitással korrigálva.

ÉV		termelés (1000 liter)
első	TÉL	560
	TAVASZ	576
	NYÁR	590
	ŐSZ	565
második	TÉL	558
	TAVASZ	581
	NYÁR	602
	ŐSZ	579
harmadik	TÉL	567
	TAVASZ	598
	NYÁR	607
	ŐSZ	600

$$\sum_{t=1}^n y_t = 6983 \quad \sum_{t=1}^n t \cdot y_t = 45\,824 \quad \sum_{t=1}^n \lg y_t = 33,176 \quad \sum_{t=1}^n t \cdot \lg y_t = 215,97$$

7. STATISZTIKAI BECSLÉSEK

7.1. Egy vasúttársaság nagysebességű járatain az utasok száma lényegében normális eloszlású. Adjunk becslést az utasok átlagos számára 90%-os konfidenciaszinten, ha 10 megvizsgált járaton az utasok száma: 360; 453; 467; 451; 487; 491; 390; 512; 488; 495.

A vonaton 480 ülőhely van. Adjunk becslést, arra, hogy az esetek hány százalékában fordul elő, hogy van ülőhely nélküli utas. (a konfidenciaszint legyen 90%)

7.2. Egy bankfiókban a sorra kerülésig eltelt idő lényegében normális eloszlású. Adjunk becslést az átlagos várakozási időre 95%-os konfidenciaszinten az alábbi minta alapján:

5; 12; 4; 7; 6; 8; 4; 3; 2; 5

Az 5 percnél hosszabb várakozást a bank vezetősége nem díjazza, és ezt igyekszik alkalmazottjai tudomására hozni. A minta alapján adjunk becslést, hogy az ügyfelek hány százalékánál haladja meg a várakozási idő az 5 percet.

7.3. Egy múzeum látogatóinak átlagéletkorát szeretnénk megbecsülni. A megkérdezett 30 ember megoszlása:

életkor	látogatók száma
10-29	5
30-49	13
50-69	12
Össz.	30

Adjunk becslést 0,95-ös konfidenciaszinten a múzeum látogatóinak átlagéletkorára és az 50 év alatti látogatók %-os arányára.

7.4. Egy étteremben felmérést készítettek arról, hogy a vendégek átlagosan mennyi időt töltenek náluk. A 40 megfigyelt vendég adatai:

Az étteremben eltöltött idő (perc)	vendégek száma
0-29	4
30-59	12
60-89	18
90-119	6
Össz.	40

Adjunk becslést 0,95-ös konfidenciaszinten az étteremben átlagosan eltöltött időre és a másfél óránál tovább maradók részarányára.

Mekkora mintára lenne szükség, ha ugyanezen a konfidenciaszinten az átlag hibáját felére akarjuk csökkenteni?

Mekkora mintára van szükség, ha az eredeti hiba mellett, de 99%-os konfidenciaszintet szeretnénk biztosítani?

7.5. Korábbi felmérések alapján valamelyik egyetemi előadás átlagos látogatottsága 98 fő, a szórás pedig 34 fő. Egy új felmérés készítését tervezik, hogy kiderüljön, van-e olyan rossz az előadás, hogy elegendő legyen 100 fős előadóban tartani. Hány előadás létszámát kell megvizsgálni, hogy az átlagos létszám becslésének hibája 10 főnél kisebb legyen 90%-os konfidenciaszinten?

7.6. Egy utazási iroda opcionálisan meghirdetett városnéző túrája minimum 35 fő esetén indul el. A csoport 70 főből áll és 40-en már nyilatkoztak, közülük 28-an mutatnak hajlandóságot városnézésre. 90%-os konfidenciaszinten kijelenthetjük-e, hogy lesz csoportos városnézés?

7.7. Egy mozi felmérést készít látogatóinak életkorával kapcsolatban. A mozilátogatók életkora tekinthető normális eloszlásúnak. A 400 fős minta eredménye:

életkor	nézők száma
0-10	7
11-20	132
21-40	157
41-60	104
össz	400

Adjunk becslést 95%-os megbízhatósággal az átlagos életkorra, szórásra és a 20 évnél idősebb nézők részarányára. Hány embert kéne megkérdezni, ha ugyanekkora hibával, de 99%-os megbízhatóságot szeretnénk?

7.8. A naponta átlagosan háztartási tevékenységgel töltött időt szeretnénk megbecsülni. A becsléshez rétegzett mintát vettünk, a rétegeképző ismérv az volt, hogy a megkérdezett nő-e vagy férfi. A nők és férfiak részaránya egyezőnek tekinthető.

	megkérdezettek száma	átlag (perc)	szórás (perc)
nő	180	74	28
férfi	120	32	19

Adjunk 90%-os megbízhatóságú becslést az átlagosan háztartási tevékenységgel töltött időre.

7.9. Egy város három kerületében 250 000, 320 000 és 180 000 lakos él. Adjunk becslést 95%-os konfidenciaszinten a naponta átlagosan utazással töltött időre az alábbi rétegzett minta alapján:

	megkérdezettek száma	átlag (perc)	szórás (perc)
1. kerület	180	75	28
2. kerület	220	54	19
3. kerület	100	43	10

7.10. Egy gyümölcslé rosttartalmára szeretnénk becslést adni 30 elemű FAE-minta alapján. A minta eredménye:

Rosttartalom (%)	elemszám
0-1,9	17
2-3,9	9
4-5,9	4
Össz.	30

Adjunk becslést 0,95-ös konfidenciaszinten az átlagos rosttartalomra. Legfeljebb a gyümölcslevek hány százaléka rendelkezhet 4%-nál nagyobb rosttartalommal 95%-os konfidenciaszinten?

8. HIPOTÉZISVIZSGÁLAT

8.1. Egy üzemben 5kg-os mosóporokat töltenek 21 gramm szórással és lényegében normális eloszlással. Az egyik gép által csomagolt mosóporok közül egy 41 elemű minta átlaga 4980 gramm, szórása 25 gramm.

5%-os szignifikanciaszinten megfelel-e a gép beállítása a szabványnak?

8.2. Egy üzemben literenként 300 mg C-vitamint adagolnak a dobozos narancslevekhez, közelítőleg normális eloszlással, 20 mg szórással. Egy szállítmányból vett 50 elemű minta átlagosan 310 mg C-vitamint tartalmazott, 22 mg szórással.

10%-os szignifikanciaszinten a szállítmány megfelelt-e a szabványnak?

8.3. Egy ásványvíz literenként 650 mg oldott ásványianyagot tartalmaz, 5 mg szórással. Az ásványianyag-tartalom eloszlása normálisnak tekinthető.

Ellenőrizzük a megadott paraméterek helyességét 10%-os szignifikanciaszinten az alábbi 6 elemű, egyenként egy literes minta alapján: 648 mg, 658 mg, 642 mg, 643 mg, 654 mg, 661 mg.

8.4. Korábbi felmérések szerint, egy múzeum látogatóinak 65%-a nő. Egy véletlenszerűen választott nap 300 látogatója közül 207 nő volt.

Ellenőrizzük a nők arányára vonatkozó állítást 10%-os szignifikanciaszinten. Mekkora az a legkisebb szignifikanciaszint, amelyen a nullhipotézis, vagyis az, hogy a látogatók 68%-a nő, még éppen elvethető?

8.5. Egy vizsgán maximum 100 pont érhető el. Az egyik vizsgán 80 hallgató vett részt, eredményeik:

pontszám	f_i
0-20	12
21-40	16
41-60	25
61-80	18
81-100	9

10%-os szignifikanciaszinten tekinthető-e a vizsgázók pontszáma egyenletes eloszlásúnak? Tekinthető-e normális eloszlásúnak?

8.6. A naponta olvasással eltöltött időről terveznek egy felmérést készíteni Németországban. A népesség korcsoportonkénti megoszlásával egybevetve tekinthető-e reprezentatívnak a felméréshez használt minta 10%-os szignifikanciaszinten?

Életkor	Népesség (%)	Minta (db)
0-14	13,5	18
15-34	25	22
35-64	41,5	27
65-	20	13
összesen	100	80

8.7. Egy légitársaság felmérést készít az utasok testsúlyával kapcsolatban. Korábbi évek adatai alapján az utasok testsúly szerinti eloszlása közelítőleg normális, 81 kg-os átlaggal és 16 kg szórással.

Ellenőrizzük az eloszlásra és a paraméterekre vonatkozó hipotéziseket az alábbi 141 elemű minta segítségével 5%-os szignifikanciaszinten.

Testtömeg (kg)	Utasok száma
0-50	13
51-70	23
71-90	56
91-110	38
111-	11
összesen	141

8.8. Egy felmérés során 400 férfit és 400 nőt vizsgáltak meg, hogy megállnak-e kocsijukkal a zebránál, ha a gyalogos a járdán várakozik. A férfiak közül 310-en, a nők közül 215-en álltak meg.

Független-e a nemtől a zebránál való megállás 5%-os szignifikanciaszinten?

Ellenőrizzük azt a hipotézist, hogy a nők a zebránál kevésbé engedik át a gyalogosokat 5%-os szignifikanciaszinten.

8.9. Egy várostól északra és nyugatra lavinafogó véderdők találhatóak. A faállomány állapotának felméréséhez mindkét véderdőben véletlenszerűen kiválasztottak 150 fenyőt, a minták eredményét tartalmazza az alábbi táblázat.

Eltér-e szignifikánsan a két véderdőben a fák átlagos életkora? Szignifikánsan egyformának tekinthető-e a két véderdő faállománya? A szignifikanciaszint legyen 5%.

Fák életkora (év)	Északi véderdő	Nyugati véderdő
0-10	13	8
11-20	28	32
21-50	67	58
51-100	31	42
101-	11	10
összesen	150	150

8.10. A nem munkával töltött aktív tevékenység (kertészkedés, sportolás, stb.) megoszlása Magyarországon és Németországban egy-egy 100 elemű minta alapján:

Nem munkával töltött aktív tevékenység időtartama naponta (perc)	HU	DE
0-50	43	10
51-100	30	35
101-150	16	27
151-200	8	20
201-250	3	8

10%-os szignifikanciaszinten a minta alapján azonosak-e a szokások a két országban?

8.11. Egy tehenészetben a tehenek tejének zsírtartalmát vizsgálták. A későbbi hasznosítás során nem kedvező, ha a zsírtartalom szórása 10%-nál nagyobb. Literenkénti 5 grammos átlagos zsírtartalommal számolva és feltételezve annak normális eloszlását, szignifikánsan eltér-e a tehenek tejének zsírtartalma a megengedett 10%-tól az alábbi 10 elemű minta alapján?

A tehen sorszáma	Zsírtartalom (gramm/liter)
17.	4,7
19.	4,9
34.	5,6
36.	4,3
37.	5,1
38.	5,4
57.	6,1
58.	5,8
63.	4,2
64.	4,2

A szignifikanciaszint legyen 5%.

8.12. Egy első osztályú almaszállítmányban az almák tömegének átlaga 110 gramm, megengedett szórása 20 gramm lehet. Ellenőrizzük 85 elemű minta alapján, hogy egy adott szállítmány megfelel-e az előírásoknak. Az almák méretének eloszlását nem ismerjük, a szignifikanciaszint legyen 10%.

alma tömege (gramm)	f_i
50-69	12
70-89	16
90-109	25
110-129	24
130-159	8

8.13. Kétféle juhajt, a német húsmerinó és az Új-Zélandi Romney gyapjájának finomságát vizsgáltuk meg. A német húsmerinó gyapjájának finomsága egy ötelemű minta alapján 26 mikron, 27 mikron, 22 mikron, 28 mikron, 30 mikron. Az Új-Zélandi Romney gyapjút egy hatelemű mintával vizsgáltuk, ennek eredménye 30 mikron, 28 mikron, 32 mikron, 33 mikron, 29 mikron, 30 mikron.

Ellenőrizzük 10%-os szignifikanciaszinten, az alábbi állításokat:

- A Romney gyapjának szálvastagsága nagyobb
- A Romney gyapjának szálvastagsága 4 mikronnal nagyobb
- A kétféle juh gyapjú-szálvastagságának szórása megegyezik

8.14. 200 fő részvételével tesztelték egy vitaminkészítmény hatékonyságát. 100-an rendszeresen szedték a készítményt, míg a másik 100 résztvevő egyáltalán nem szedett semmit, vagy másfajta vitaminokat szedett. Az évente betegség miatt kieső munkanapok számát hasonlították össze a két csoportban, ezek eloszlását normális eloszlásúnak tekinthetjük. 5%-os szignifikanciaszinten mi mondható az alábbi állításokról?

csoportok	Betegség miatt kieső munkanapok	
	átlaga	szórása
Szedték a készítményt	7,2	3,7
Nem szedték a készítményt	7,8	3,4

Megegyezik-e a két csoportban a kieső munkanapok átlaga és szórása?

Szignifikánsan eltér-e a betegség miatt kieső munkanapok száma a két csoportban?

8.15. Egy gyártósor gumicukrokat tölt zacskókba. A zacskóknak azonos arányban kell tartalmaznia kék, piros, zöld, sárga és fehér színű gumicukrokat. Ellenőrizzük 4 zacskó megvizsgálásával egy 80 elemű minta alapján, hogy 5%-os szignifikanciaszinten, hogy a gép megfelel-e a szabványnak.

szín	db
Kék	24
Piros	16
Zöld	12
Sárga	18
Fehér	10

8.16. A naponta utazással eltöltött időt vizsgálták középiskolások és egyetemisták körében. A középiskolások utazással töltött ideje egy 60 elemű minta alapján naponta átlag 83 perc, a szórás 17 perc. Ugyanez az egyetemistáknál a következőképpen alakult:

Utazással töltött idő (perc)	Válaszolók száma
-50	12
51-100	36
101-200	24
201-	8

5%-os szignifikanciaszinten megegyezik-e a két csoportban a naponta utazással töltött idő átlaga és szórása? Szignifikánsan eltér-e az utazással eltöltött idő a két csoportban?

9. REGRESSZIÓSZÁMÍTÁS

9.1. Néhány ország közép fokú iskolai képzésének egy diákra jutó oktatási ráfordítása illetve az éves egy főre jutó GDP adatai láthatók az alábbi táblázatban. Adjuk meg a lineáris regressziós modellt, a reziduális szórást, határozzuk meg a modell magyarázó erejét.

ország		x	y
		GDP/fő (EUR)	Oktatási ráfordítás (Közép fokú képzés diák/EUR)
Ausztria	AT	28 978	76 900
Belgium	BE	30 349	61 000
Csehország	CZ	15 216	33 800
Franciaország	FR	26 656	57 600
Görögország	GR	17 941	59 200
Hollandia	NL	28 669	61 500
Lengyelország	PL	10 135	30 700
Magyarország	HU	13 767	33 000
Németország	DE	28 232	65 300
Svájc	CH	31 987	60 400

9.2. Egy strand forgalmának alakulása a napi középhőmérséklettől függően 12 megfigyelt nap alapján az alábbi volt:

nap	napi közép- hőmérséklet (°C)	forgalom (fő)
1.	22	765
2.	23	1572
3.	18	510
4.	25	1967
5.	22	1142
6.	16	576
7.	24	986
8.	20	1216
9.	24	1267
10.	26	1686
11.	19	981
12.	20	1412

Adjuk meg a lineáris regresszió egyenletét, adjuk meg a korrelációs és a determinációs együtthatót és döntsük el, hogy a lineáris vagy a hatványkitevős regresszió illeszkedik-e jobban, ha ismeretes, hogy

$$\sum d^2x = 100,91 \quad \sum d^2y = 2\,155\,847 \quad \sum dx \cdot dy = 10\,894,67 \quad \hat{y} = 1,43 \cdot x^{2,17}$$

9.3. Az alábbi táblázat néhány ország egy főre jutó GDP-jét és a nők életkorát tartalmazza első házasságkötésük idején. Készítsünk lineáris regressziót, ahol a magyarázó változó az egy főre jutó GDP. Értelmezzük a modell paramétereit, készítsünk varianciaanalízis táblázatot, adjuk meg a modell magyarázó erejét!

ország		GDP/fő (EUR)	Nők életkora házasságkötéskor
Ausztria	AT	28 978	26,6
Belgium	BE	30 349	29,8
Csehország	CZ	15 216	28,9
Franciaország	FR	26 656	31,6
Görögország	GR	17 941	26,9
Hollandia	NL	28 669	26,9
Lengyelország	PL	10 135	25,3
Magyarország	HU	13 767	29,7
Németország	DE	28 232	31
Svájc	CH	31 987	29,4

$$\sum d^2x = 579\,956\,336 \quad \sum dx \cdot dy = 56\,484$$

9.4. Néhány ország adatai alapján vizsgáljuk meg az átlagos iskolázottsági szint és a születéskor várható élettartam közti kapcsolatot. Adjunk meg a lineáris és az exponenciális regressziós modellt, amiben magyarázó változó az átlagos iskolázottsági szint. Melyik modell illeszkedik jobban?

	Átlagos iskolázottsági szint (év)	Születéskor várható élettartam (év)
1.	12,6	81,1
2.	12,4	78,5
3.	11,6	75,4
4.	10,4	74
5.	4,4	65,4
átlag	10,3	74,9

9.5. Egy cégnél 30 alkalmazottat vizsgáltak meg, hogy miként magyarázza az életkor, illetve az, hogy az illető férfi-e vagy nő (férfi=0, nő=1) a fizetés nagyságát. A kapott regressziós modell a havi fizetés nagyságát ezer forintban adja meg, ahol x_1 jelenti az életkort és x_2 jelenti azt, hogy az illető férfi-e vagy nő.

$$\hat{y} = 64 + 7,6x_1 - 16,7x_2 \quad s_{\hat{\beta}_1} = 4,2 \quad s_{\hat{\beta}_2} = 10,83 \quad SSE = 81,2 \quad SST = 105,7$$

Adjuk meg a modell paramétereinek jelentését. Szignifikánsnak tekinthető-e modell alapján az életkor, illetve a nem, az alkalmazott fizetése szempontjából 10%-os szignifikanciaszinten? Teszteljük a teljes modellt 10%-os szignifikanciaszint mellett.

9.6. Egy strand forgalmának modellezésére két magyarázó változót használunk, a napi középhőmérsékletet (x_1) illetve azt, hogy hétvége van-e vagy sem ($x_2=0$ ha nincs hétvége és $x_2=1$ ha igen). Egy 12 megfigyelés alapján készített modellről az alábbiakat tudjuk:

$$\hat{y} = 396 + 12,6x_1 + 18x_2 \quad s_{\hat{\beta}_1} = 2,19 \quad s_{\hat{\beta}_2} = 38,15$$

$$R = \begin{pmatrix} 1 & & \\ 0,92 & 1 & \\ -0,57 & -0,67 & 1 \end{pmatrix}$$

Adjuk meg a lineáris regressziós modell paramétereinek jelentését. Szignifikánsnak tekinthető-e modell alapján a napi középhőmérséklet a strand forgalmának szempontjából 10%-os szignifikanciaszinten? Adjuk meg a forgalom és a hőmérséklet kapcsolatát leíró parciális korrelációs együttható értékét. Adjuk meg a többszörös determinációs hányados értékét.

mateking.hu