

V. modul: Idősorok elemzése

11. lecke Analitikus trendszámítás

Ha a vizsgált jelenség tartós irányzatát az idő függvényében valamilyen regressziós függvény-nel határozzuk meg, akkor analitikus trendszámításról beszélünk. Az analitikus trendszámítás tehát a regresszió-számítás egy speciális esete, amennyiben az idősorban bekövetkezett változásokat az időtényező (t) függvényében vizsgáljuk.

Hasonlóan a regresszió-számításhoz, a trendszámításnál is először ábrázolni kell az adatokat, és kiválasztani a függvény típusát.

11.1. Lineáris trend

Ha az adatokra az egyenletes változás jellemző, akkor lineáris trenddel határozzuk meg az alapirányzat értékeit.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 * t$$

Hasonlóan a regresszió-számításhoz, itt is a legkisebb négyzetek módszerét alkalmazzuk, a normálegyenletek a következők:

$$\sum y_i = n * \hat{\beta}_0 + \hat{\beta}_1 * \sum t_i$$

$$\sum t_i * y_i = \hat{\beta}_0 * \sum t_i + \hat{\beta}_1 * \sum t_i^2$$

A megoldáshoz kódolnunk kell az idősor adatait. Ez többféleképpen lehetségesféleképpen történhet:

- Ha az idősort $t=1,2,3,\dots,n$ kódoljuk:

év	t_i
2009	1
2010	2
2011	3
2012	4
2013	5
2014	6
2015	7
2016	8

A normálegyenleteket átalakítva megkapjuk a trendfüggvény együtthatóit:

$$\hat{\beta}_1 = \frac{\sum t_i * y_i - \frac{(\sum t_i) * (\sum y_i)}{n}}{\sum t_i^2 - \frac{(\sum t_i)^2}{n}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 * \frac{\sum t_i}{n}$$

- Ha az idősort $\Sigma t=0$ módon kódoljuk, akkor különbség van a páros és páratlan számú idősor esetében.
- Páratlan tagszámú idősor:

év	t
2009	-3
2010	-2
2011	-1
2012	0
2013	1
2014	2
2015	3

- Páros tagszámú idősor esetén:

év	t	t
2009	-3,5	-7
2010	-2,5	-5
2011	-1,5	-3
2012	-0,5	-1
2013	0,5	1
2014	1,5	3
2015	2,5	5
2016	3,5	7

Ebben az esetben a normálegyenletek egyszerűbbek, és a trendfüggvény együtthatóit az alábbi képletekkel kapjuk meg:

$$\sum y_i = n * \hat{\beta}_0$$

$$\sum t_i * y_i = \hat{\beta}_1 * \sum t_i^2$$

$$\hat{\beta}_0 = \frac{\sum y_i}{n} = \bar{y}$$

$$\hat{\beta}_1 = \frac{\sum t_i * y_i}{\sum t_i^2}$$

A $\hat{\beta}_1$ paraméter az időegységenkénti átlagos abszolút változás mértéke, előjelétől függően növekedést vagy csökkenést jelez.

Ha $\sum t=0$, és az időpontok száma páros, akkor $2 * \hat{\beta}_1$ az időegységenkénti átlagos abszolút változás mértéke.

A $\hat{\beta}_0$ paraméter az alapirányzat értéke a $t=0$ időpontban:

- ha $t=1, \dots, n$, akkor a vizsgálatba bevont időpontot megelőző időpont trend szerinti értéke.
- Ha $\sum t=0$, és páratlan az idősornál választott tagszám, akkor a középső időpont alapirányzata és egyben a vizsgált idősor adatainak számtani átlaga.
- Ha $\sum t=0$, és páros az idősornál választott tagszám, és nincs $t=0$, akkor az idősor adatainak számtani átlaga.

Ezután a regresszió számításához hasonlóan, kiszámoljuk a reziduális szórásnégyzetet.

$$s_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

és a reziduális szórás mutatószámát:

$$V_e = \frac{s_e}{\bar{y}} * 100$$

Ez a valós értékektől való átlagos eltérését mutatja a trendértékeknek.

1. Bemutató feladat

A villamosenergia-termelés Magyarországon.

	millió kWh	t_i	t^2	$t * y$	y_t	$(y - y_t)^2$
2006	29580	-5	25	-147900	28451,18	1274234,59
2007	28365	-4	16	-113460	29461,20	1201650,06
2008	29932	-3	9	-89796	30471,22	290753,89

2009	31238	-2	4	-62476	31481,23	59162,78
2010	32630	-1	1	-32630	32491,25	19251,01
2011	33928	0	0	0	33501,27	182098,49
2012	34978	1	1	34978	34511,29	217820,09
2013	35305	2	4	70610	35521,31	46788,29
2014	37023	3	9	111069	36531,32	241745,29
2015	36968	4	16	147872	37541,34	328721,05
2016	38567	5	25	192835	38551,36	244,61
Összesen	368514	0	110	111102	368513,97	3862470,15
átlag	33501,27					

$$\hat{\beta}_1 = 1010,018$$

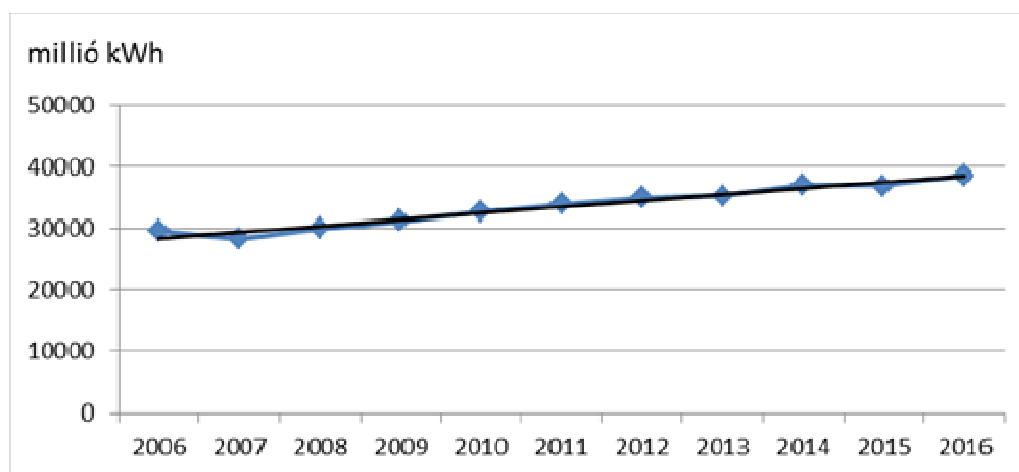
$$\hat{\beta}_0 = 33501,27$$

Az y_t -értékeket úgy kapjuk meg, hogy a trendegyenletbe rendre behelyettesítjük a t -értékeket.

$$y_t = 33501,27 + 1010,018 \cdot (-5) = 28451,18$$

$$y_t = 33501,27 + 1010,018 \cdot (-4) = 29461,20$$

stb.



$$s_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n} = \frac{3862470,15}{11} = 351133,65$$

$$s_e = 592,565$$

$$V_e = \frac{s_e}{\bar{y}} * 100 = \frac{592,565}{33501,27} * 100 = 1,77 \%$$

A függvény szerinti villamosenergia-temelés 1,77%-os hibával illeszkednek a tényleges adatokhoz.

Nézzük meg, ha másmódon kódoljuk az időt, mi lesz a trendegyenlet!

	millió kWh	t	t ²	t*y	yt	y-yt	(y-yt) ²
2006	29580	1	1	29580	28451,17	1128,83	1274257,17
2007	28365	2	4	56730	29461,19	-1096,19	1201632,52
2008	29932	3	9	89796	30471,21	-539,21	290747,424
2009	31238	4	16	124952	31481,23	-243,23	59160,8329
2010	32630	5	25	163150	32491,25	138,75	19251,5625
2011	33928	6	36	203568	33501,27	426,73	182098,493
2012	34978	7	49	244846	34511,29	466,71	217818,224
2013	35305	8	64	282440	35521,31	-216,31	46790,0161
2014	37023	9	81	333207	36531,33	491,67	241739,389
2015	36968	10	100	369680	37541,35	-573,35	328730,223
2016	38567	11	121	424237	38551,37	15,63	244,2969
Összesen	368514	66	506	2322186	368513,97		3862470,15
átlag	33501,27						

$$\Sigma y_t = n * b_0 + b_1 * \Sigma t$$

$$\Sigma t * y_t = b_0 * \Sigma t + b_1 * \Sigma t^2$$

$$368514 = 11 * b_0 + b_1 * 66$$

$$2322186 = b_0 * 66 + b_1 * 506$$

$$b_1 = 1010,02$$

$$b_0 = 27441,15$$

$$y_t = 27441,15 + 1010,02 * t$$

11.2. Exponenciális trend

A társadalmi-gazdasági folyamatok változó környezetben nem mindig mutatnak lineáris tendenciát. Ha az időegységenkénti relatív változás ingadozik egy állandó körül, akkor a tartós irányzatot exponenciális trenddel fejezzük ki. Az exponenciális trendegyenletet logaritmus segítségével visszavezetjük lineárisra:

$$\hat{y} = \hat{\beta}_0 * \hat{\beta}_1^t$$

$$\lg \hat{y} = \lg \hat{\beta}_0 + t * \hat{\beta}_1$$

$$Y = B_0 + t * B_1$$

A normálegyenlet:

$$\Sigma Y_i = n * B_0 + B_1 * \Sigma t_i$$

$$\Sigma t_i * Y_i = B_0 * \Sigma t_i + B_1 * \Sigma t_i^2$$

A $\Sigma t=0$ esetében a normálegyenletek leegyszerűsödnek:

$$\sum Y_i = n * B_0$$

$$\sum t_i * Y_i = B_1 * \sum t_i^2$$

$$B_0 = \frac{\sum Y_i}{n}$$

$$B_1 = \frac{\sum t_i * Y_i}{\sum t_i^2}$$

$$\hat{\beta}_0 = 10^{B_0}$$

$$\hat{\beta}_1 = 10^{B_1}$$

A β_1 paraméter az időegységenkénti relatív változást mutatja.

1. Bemutató feladat

Egy üdülőövezet vendégeinek száma látható a táblázatban, határozzuk meg az exponenciális trendet!

Exponenciális trend esetében először logaritmizálni kell az y_i értékeket. A továbbiakban ezekkel az értékekkel számolunk.

	efő	t	lgy	t*lgy	t ²	y _t	y-y _t	(y-y _t) ²
2006	136	-5	2,133539	-10,6677	25	119,7789	-16,2211	263,1228
2007	152	-4	2,181844	-8,72737	16	134,9909	-17,0091	289,3107
2008	157	-3	2,1959	-6,5877	9	152,1347	-4,86529	23,67109
2009	166	-2	2,220108	-4,44022	4	171,4558	5,455813	29,76589
2010	183	-1	2,262451	-2,26245	1	193,2307	10,2307	104,6672
2011	188	0	2,274158	0	0	217,771	29,771	886,3124
2012	185	1	2,267172	2,267172	1	245,4279	60,42792	3651,533
2013	225	2	2,352183	4,704365	4	276,5973	51,59726	2662,277
2014	341	3	2,532754	7,598263	9	311,7251	-29,2749	857,0189
2015	440	4	2,643453	10,57381	16	351,3142	-88,6858	7865,17
2016	450	5	2,653213	13,26606	25	395,9311	-54,0689	2923,445
	2623	0	25,71677	5,724238	110			19556,3
	238,4545	0	2,337888					

$$B_0 = \frac{\sum Y_i}{n} = \frac{25,71677}{11} = 2,338$$

$$B_1 = \frac{\sum t_i * Y_i}{\sum t_i^2} = \frac{5,724238}{110} = 0,052039$$

$$\hat{\beta}_0 = 10^{B_0} = 10^{2,338} = 217,771$$

$$\hat{\beta}_1 = 10^{B_1} = 10^{0,052039} = 1,127$$

$$s_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n} = \frac{19556,3}{11} = 1777,8455$$

$$s_e = 42,16$$

$$V_e = \frac{s_e}{\bar{y}} * 100 = \frac{42,16}{238,4545} * 100 = 17,68 \%$$

A függvény szerinti létszámok 17,68%-os hibával illeszkednek a tényleges adatokhoz.

11.3. Választás a trendegyenletek közül

Hasonlóan a regresszióanalízishez, a trendszámításnál is sok esetben nehéz eldönteni, hogyan melyik függvényt használjuk az elemzéshez. A kiválasztáshoz a legkisebb négyzetek módszerét használjuk. Az a függvény illeszkedik a legjobban az idősor adataihoz, amelyiknél

$$\sum (y_i - \hat{y}_i)^2 \text{ a legkisebb.}$$

11.4. Extrapoláció

A trendegyenlet meghatározásával előrejelzést (extrapoláció) végezhetünk. Fontos, hogy előrejelzést csak abban az esetben végezhetünk, ha:

- Kellően hosszú idősből határoztuk meg a trendegyenletet.
- Illetve, ha várhatóan nem következik be olyan változás, amely a vizsgált jelenséget nagymértékben megváltoztatja.