# Idősor elemzési modellek gyakorlati alkalmazása újszülött malac aEEG adatain

Statistical Analysis Software Enterprise Guide 4.3 OnDemand és IBM SPSS Statistics Version 21 felhasználásával

#### László Anna

#### PhD hallgató

Szegedi Tudományegyetem, Általános Orvostudományi Kar, Interdiszciplináris Orvostudományok Doktori Iskola

2013. augusztus 31.

#### Tartalomjegyzék

1	E	Bevezetés
2	k	Kísérlet asphyxiás malacmodellel
3	A	Amplitúdó integrált EEG
4	S	Szűrés (Filtrálás)5
5	A	Adatok beolvasása és megjelenítése SAS-sal 6
	5.1	Adatimport6
	5.2	Mért adatok grafikus megjelenítése az idő függvényében7
6	I	dő alapú jelfeldolgozás SAS szoftver felhasználásával 11
	6.1	Artefaktumok kiszűrése 11
	6.2	Egyenirányítás (Rektifikálás)14
	6.3	Normalizálás
	6.4	Átlagolások 15
7	I	dősoros elemzés IBM SPSS Statistics felhasználásával19
	7.1	Adatok beolvasása
	7.2	Adatok szűrése 23
	7.3	Megjelenítés

7	7.4	Arte	efaktumok kiszűrése	26
7	7.5	Nor	rmalitás ellenőrzés	29
7	7.6	Aut	tokorreláció	32
7	7.7	Gör	rbeillesztés	35
7	7.8	Szez	zonális dekompozíció, mint determinisztikus matematikai modell	39
	7.8.	1	Multiplikatív modell	41
	7.8.	2	Additív modell	51
7	7.9	Exp	oonenciális simítás	59
7	7.10	ARII	IMA modellillesztés	53
7	7.11	Spe	ektrál elemzés	57
8	Kösz	zönet	tnyilvánítás	70
9	Irod	lalom	njegyzék	71
10	Т	ámog	gatás	72

## 1 Bevezetés

Jelen dokumentumban általánosan az időben zajló biológiai folyamatok adatelemzéséhez találunk gyakorlati alkalmazásokat SAS és IBM SPSS adatelemző szoftverek felhasználásával. A biológiai jelek közül most EEG, konkrétan aEEG mért értékekkel dolgozunk, de hasonlóan a kidolgozott SAS valamint SPSS kódrészletek aktualizálhatók bármely más jel elemzéséhez. Hiszen, ahogy Hajtman Béla írta: a statisztikai módszerek nem tesznek különbséget az adatok közt aszerint, hogy honnan származnak, lehetnek orvosi, pszichológiai, műszaki, társadalmi vagy akár gazdasági eredetűek is (1).

A 6.4 fejezetig már az előző negyedéves teljesítés részeként leadott SAS elemzéseket tartalmazza a dokumentum. Ennek kiegészítését olvashatjuk a további fejezetekben, elsősorban az idősoros elemzés alap módszereit ismerhetjük meg az IBM SPSS szoftver felhasználásával, a kapott eredmények értelmezésével.

## 2 Kísérlet asphyxiás malacmodellel

Érett újszülöttekben előfordul szövődményként az aszfixia (légzési elégtelenség), amely lehet súlyos maradandó idegrendszeri károsodás, de akár halál okozója is. Előfordulhat enyhébb hatás is, de előbbi veszélyek lehetősége miatt érdemes nagy hangsúlyt fektetni a terület vizsgálatára. Munkacsoportunk az aszfixia hatásait vizsgálta.

Az újszülött malac az érett humán újszülött egyik leggyakrabban használt nagy állatmodellje (megkülönböztetésül a rágcsálómodellektől) (2, 3). A tudományos kísérlet során újszülött malacok aszfixiás állapota során fellépő agykárosodás és az ezt követő regenerálódási folyamatok vizsgálata történt. A humán csecsemő születésekor felmerülő cerebrovaszkuláris fiziológiai folyamatokat modellezte a munkacsoport. Az eredmények nemrég kerültek publikálásra (4).

Az aszfixiát a lélegeztetés felfüggesztésével hozták létre. Ilyenkor *hipoxia* (oxigénhiány) és *hiperkapnia* alakul ki (szén-dioxid felhalmozódik a szervezetben). A szív hipoxiája miatt a pumpafunkció károsodik, és 2-3 perc után már a vérellátás jelentős csökkenése, *iszkémia* (vérellátás zavara) is kialakul, főleg azért, mert a keringést fenntartó perfúziós nyomás (artériás – vénás vérnyomás) a szívpumpa kialakuló elégtelensége miatt csökken. Az aszfixia szövődményeként kialakult kórképet *hipoxiás-iszkémiás enkefalopátiá*nak (HIE) is nevezik (az enkefalopátia az agy különböző eredetű globális működészavarait jelenti).

A vizsgálat célja elsősorban azon kérdés megválaszolása volt, hogy *egy aszfixiás periódus károsítja-e az agyi keringésszabályozásban fontos szerepet játszó mechanizmusokat az aszfixiás periódus után 24 órával* (azt már jól tudta a munkacsoport, hogy 1 órával később van károsodás, ami rendeződni látszik 4 órára (5, 6)).

Második kérdés - ami az első kérdés pozitív válasza után merülhetett fel, hogy *a hidrogén véd-e a károsodás kialakulásától*, mert egy korábbi tanulmány szerint az akut károsodás ellen már védett (7).

Újszülött malacokból három csoport került be az egész napos kísérletbe ezek egy kontroll, egy aszfixiás és egy hidrogénnel lélegeztetett aszfixiás csoportok voltak. Mindhárom csoportban 9-9 alany volt, összesen tehát n=27 eset állt rendelkezésre. A kísérlet során legelőször fel lett véve egy alap amplitúdó integrált EEG (aEEG) jel kb. 10 percig mindhárom csoportban, majd 8 percen keresztüli lélegeztetés elzárással aszfixiás állapot lett előidézve két csoportnál. Ennek eredményeképp kb. 1,5 perc után kisimult az aEEG görbe, de az állatok szíve nem állt le. Az így előálló hipoxiás-iszkémiás állapotban az egyik csoportot az aszfixiás sokk után 4 órán keresztül 2,1% hidrogén tartalmú szobalevegővel lélegezteték újra (a reventilláció során adott levegőben a szokásos normál 21% oxigén mellett a nitrogénből volt kicsit kevesebb a hidrogén javára).

Az aszfixia alatti aEEG is rögzítésre került. Az aszfixiás stresszt követően további 22 órán keresztül mérte a munkacsoport az aEEG értékeket, óránként kb. 10 percig, 10 Hz-en (vagyis 0,1 szekundumonként van adatunk, minden óra elején kb. 6000 db). Az adatok az eddigiekben nem kvantitatívan lettek elemezve, hanem a felvételek pontozásra kerültek. Az aEEG-n a kérgi elektromos aktivitás regenerációját lehet nagyjából nyomon követni, de idősoros elemzés, mélyebb összefüggések keresése még nem történt. Jelen dokumentumnak elsősorban ezen információk kinyerésének elindítása a célja, különböző módszerek alkalmazásával és ismertetésével.

Emellett egyéb akut mérések is történtek a hipoxiás-iszkémiás folyamatokból történő regeneráció állapotának felmérésére. Ezek a rögzített hemodinamikai paraméterek, mint a *testhőmérséklet*, *szívfrekvencia, szisztolés vérnyomás* vagy *szaturáció* (SpO<sub>2</sub>) minden órában, illetve *vérgáz* (pCO<sub>2</sub>), *pH* és *glükóz* négy óránként a két aszfixiás csoport között nem mutatott szignifikáns különbséget (kétszempontos ismételt méréses ANOVA, SNK post hoc próba). Emellett EKG jel is rögzítésre került.

A 22 óra után egy kb. kétórás beavatkozással a koponyába egy zárt ablak került implantációra, mellyel helyre lehet állítani a koponyaűr zártságát, így nyomásviszonyait. Az agyat burkok veszik körül, ezekből a külsőket eltávolították (kemény: *dura mater*). A harmadik burok szorosan az agykéreg felszínéhez tapad (*pia mater*). Ebben a legbelső agyhártyában ágaznak szét az erek, innen fúródnak az agykéreg állományába (*parenchyma*). Itt a pia materben az agykéreg felszínén futó, 300 mikrométernél kisebb belső átmérőjű verőerek (*pial arterioles*, arteriolák: verőerecskék) közül egynek nyomon követték az átmérőváltozásait az ablakon át egy operációs mikroszkóp segítségével. Ezek az erek már jelentősen befolyásolják az agykérgen átáramló vér mennyiségét, a szabályozásban fontos szerepet töltenek be.

A fontos méréseket a fent említett beavatkozás után, az aszfixiát követő 24-edik órában végezte a munkacsoport, amikor is az agyi ér reaktivitását mérték 4 paraméter alapján: *hiperkapnia* (vér magas széndioxid szintje), *NMDA*, *noradrenalin* (NA, vagy más néven norepinefrin – NE, egyfajta

katekolamin, ami hormonként és neurotranszmitterként is működhet), *SNP* (sodium-nitroprusside: egy értágító – vazodilatátor - hatású só). Hisztopatológiai vizsgálat is történt.

Az összes mérést a 24 órás túlélés alatt azért végezték, hogy legyen adat arról, hogy az érválaszok közötti esetleges különbségeket **nem** az egy nap túlélés alatti fiziológiai különbségek hozhatják létre. Különösen fontos például a testmag hőmérséklete (hipotermia protektív), és a vérnyomás.

A pial arteriolákon kimutatták, hogy aszfixia után egy nappal is jelentősen megváltozott a reaktivitásuk a szén-dioxiddal és az ún. NMDA-val szemben.

## 3 Amplitúdó integrált EEG

Az Elektroencefalográfia (EEG) módszerével az agy elektromos tevékenységét tudjuk időben követni (8). A mérőműszerről elvezetett jel, az ún. elektroenkefalogram egy több összetevőből álló nemperiodikus görbével írható le. A görbe karakterisztikáját különféle módszerekkel elemzik, mely során a háttérben zajló specifikus szenzoros, motoros és kognitív folyamatokat próbálják feltárni, jellemezni (9).

A nyers jel amplitúdójának értelmezési tartománya kb. 100 µV-ig terjed.

A klasszikus frekvenciatartománya 0,5-30 Hz között mozog (delta, theta, alpha, beta hullámok), de felmehet akár 70-80 Hz-re is (gamma hullámok). Ez utóbbiak újszülöttben jellemzően nem fordulnak elő, heveny kognitív erőfeszítésnél szoktak ilyen sűrű jelet rögzíteni.

Az EEG jelből rektifikálást és integrálást követően kapunk amplitúdó integrált EEG (aEEG) jelet.

## 4 Szűrés (Filtrálás)

Nyers jelből (nem feltétlenül EEG, lehet EMG, vagy egyéb mért jel is) a nemkívánatos zaj kiszűrése a mérőműszerbe beépített integrált szűrővel történik (10). Különböző szűrők léteznek, ilyenek például a következők:

- *Alul áteresztő szűrő* (LPF: low-pass filter, high-cut filter): Adott Hz fölötti nemkívánatos frekvenciák kiszűrésére alkalmas. Ez szűri ki a váltakozóáramú hálózati zavart. **Ez a torzulás később nem szűrhető ki.**
- *Felül áteresztő szűrő* (HPF: high-pass filter, low-cut filter): hogy kiszűrjük a membrán potenciálnak köszönhető egyenáram szüneteket, ill., hogy minimalizáljuk az elektródának a bőr felszínén való mozgásából következő alacsony frekvenciás interferenciát. Ezzel szűrhetjük ki az alapvonal ingadozásait.
- *Sávszűrő* (bandpass): Egyben alul és felül áteresztő szűrő is.
- *Sávzáró:* Lassan és gyorsan váltakozó zaj jelek kiszűrése

• Hálózati zajszűrő: 50 Hz körüli lyukszűrő.

## 5 Adatok beolvasása és megjelenítése SAS-sal

#### 5.1 ADATIMPORT

Első körben csak egy esethez tartozó adatokat olvassuk be, ez a kísérletből az egyik hidrogénnel lélegeztetett aszfixiás malac volt.

Az adatbeolvasásra és a későbbi adatfeldolgozásra- illetve elemzésre a SAS Enterprise Guide 4.3 OnDemand alkalmazást használtam, de a SAS Base környezetben is futtatható kódok megtalálhatóak jelen dokumentumban.

A következő kódrészlet egy Excel file-ból az adatbeolvasásra előkészített szerkezetű munkalapjának adatimportját végzi el. Ez az adatelőkészítés annyit jelent, hogy az adatok úgy kerültek strukturálásra, hogy az Excel munkalap - melynek neve "ForSAS" (lásd a kódban) - első sorában az ékezettől, szóköztől és mindenféle speciális karaktertől mentes érthető (beszédes, de rövid, használható) változónevek szerepelnek, az alatta lévő sorokban pedig az egyes időpillanatokban (tizedmásodperc) mért értékek szerepelnek. Amennyiben hiányzó érték van, annak helyét üres cella jelzi.

SAS Enterprise Guide OnDemand alkalmazásban a **File** menü **Import data** almenüjében megjelenő varázslóval végig is kattinthatjuk az adatimport paraméterezését. Az adatbeolvasás eredményét az Enterprise Guide Process Flow munkaterületén az alábbi node-ok (csomópont) jelzik.



Közvetlenül az adatimport után rögtön megnyitja a program az adatállományt, mely a későbbiekben is megtekinthető a Data Imported csomópontra kattintva.

Import	: Data (m767_1_	AH2.xls[fc	orSAS]) 🔻																						
( 🛄 C	Code ( Code Code Code Code Code Code Code Code																								
<b>\$5</b> [	💐 Modify Task	🖏 Filter	and Sort	🗒 Qu	ery Builder	Data	- Describ	ie 🕶 Graj	oh + Ar	halyze ·	- Export	- Sen	d To 🕶	000											
	😡 time 😡	) alap 🔞	asph	🌖 h1 🚺	🕖 h2 😡	) h3 🧕	) h4 迿	h5 🔞	h6 迿	h7 🧕	) h8 迿	h9 😡	h10 🔞	) h11 😡	h12 🔞	) h13 😡	) h14 😡	h15 😡	h16 🧕	🌒 h17 😡	) h18 📵	h19 🔞	) h20 👩	🌖 h21 🔞	h22
1	1	17	9	1	0	1	9	2	1	3	5	5	7	6	7	24	23	19	6	18	7	3	4	16	16
2	2	15	7	1	1	2	2	4	2	6	1	8	18	10	6	22	16	29	8	3	22	10	5	11	13
3	3	4	16	1	1	0	5	4	1	9	2	5	19	21	3	7	37	49	14	11	17	11	16	7	26
4	4	9	3	1	1	0	5	5	1	8	3	6	7	14	22	10	33	25	7	12	31	9	25	7	6
5	5	17	11	1	0	1	1	4	0	6	6	6	9	11	23	2	27	36	26	6	35	13	18	10	3
6	6	8	17	1	0	2	3	6	4	3	3 15	5	18	7	17	10	6	4	6	15	8	14	50	4	3
7	7	4	11	1	1	1	2	0	2	1	7	4	25	3	2	16	6	26	8	19	3	14	21	9	8
8	8	5	6	2	0	1	5	2	2	4	6	15	23	6	6	8	13	25	19	14	8	9	12	30	6
9	9	2	8	1	1	1	1	1	2	2	9	7	35	3	3	4	4	22	25	27	6	4	13	39	27
10	10	6	5	1	1	1	2	6	2	5	13	8	43	4	6	23	17	44	16	6	8	5	4	31	14
11	11	3	1	2	1	1	2	1	1	8	5	8	20	6	4	31	24	13	20	11	13	18	11	21	7
12	12	2	4	1	2	1	2	4	1	6	9	9	18	7	6	43	30	5	11	6	7	4	27	32	18
13	13	2	12	1	0	2	1	6	4	10	26	6	39	17	6	45	21	17	9	17	14	14	28	8	7
14	14	6	9	1	2	1	1	4	10	8	16	14	15	5	4	20	22	24	4	34	14	2	29	8	15
15	15	4	8	0	0	1	0	3	8	6	9	11	11	4	15	69	36	5	11	6	12	12	21	21	21
16	16	8	7	1	1	2	1	6	5	2	19	15	16	6	21	22	45	10	17	20	12	12	10	14	8
17	17	3	4	1	n	2	1	4	3	8	10	7	6	8	19	38	38	20	23	28	7	45	6	44	2

#### 5.2 MÉRT ADATOK GRAFIKUS MEGJELENÍTÉSE AZ IDŐ FÜGGVÉNYÉBEN

A Process Flow-n egy Program node-ba (jobb egérgombbal felvehetünk egy új programot a New  $\rightarrow$  Program funkcióval) beírhatóak az alábbi kódrészletek, melyekkel a SAS GRAPH felhasználásával megjeleníthetjük az egyes mért értékeket időben. Az egész kódsort egy GOPTIONS keretbe célszerű foglalni, hogy a korábbi és későbbi grafikus beállításoktól elkülönüljön.

Azt, hogy melyik mért értéket (jelet) szeretnénk ábrázolni az idő függvényében, a PROC GPLOT eljárás PLOT utasításában adjuk meg. Jelen példában az alapjelet tüntetjük föl ("alap" a változó neve is, ez az a kb. 10 percig mért jel, amit az aszfixiás sokk előidézése előtt mértek). A következő néhány hasonló kódrészletben látható, hogy ez a változó van egyrészt kicserélve az aktuális mért jel változójára (asph, h1, h2). Az aktuális mért jeltől függ, hogy a diagram címének mit adunk. Ezt a PROC GPLOT előtt a TITLE sorában az idézőjelek között módosíthatjuk az adott tartalom szerint. Ha a mérések száma meghaladja a jelen példa szerint kirajzolt 7000 db-ot, akkor az AXIS1 tengelybeállításoknál az ORDER opcióban hasonló 1000-es lépésköznél az 1000 adott számú többszörösére módosítsuk a vízszintes tengelyen látható érték intervallumot (idő), hogy minden mért adat látszódjon. Ha egyéb formai beállításokon (pl. betűtípusok, méretek, színek, stb.) nem szeretnénk változtatni, akkor ennyi elegendő hasonló frekvencián mért aEEG jel esetén (módosítandók tehát a szürkével jelölt részek). A fent látott adatállomány többi változójára ennek mintájára készíthetjük el az ábrázolást.

Más jel és más frekvencia esetén a tengelyek feliratait (AXIS LABEL) is módosítanunk kell, illetve természetesen magát az adatállományt (PROC GPLOT DATA opciójában) (11).

```
/* mért amplitúdó integrált EEG (aEEG) értékek szemléltetése */
goptions reset=all border ctext=black ftext="Arial" htext=10 pt;
axis1 minor=none label=("Time (0.1 sec)") offset=(5, 5)
    value=(rotate=0 angle=0) order=(0 to 7000 by 1000)
axis2 minor=none label=(angle=90 "aEEG");
symbol1 line=1 ci=black cv=black value=Dot interpol=join width=1 height=1 pt;
title height=14pt 'Alap amplitúdó integrált EEG értékek';
proc gplot data=m767_1_ah2;
    plot alap*time
    / haxis=axis1
    vaxis=axis2
    cframe=white
    caxis=black;
run; quit;
goptions reset=all;
```

A következő három kódrészlet az előző mintájára készült, az aszfixiás, az első és a második órában mért aEEG adatokat szemlélteti vonaldiagramon. A módosítások itt is szürkítve vannak. Eszerint a többi órában mért jeleket is lehet ábrázolni.

```
goptions reset=all border ctext=black ftext="Arial" htext=10 pt;
axis1 minor=none label=("Time (0.1 sec)") offset=(5, 5)
      value=(rotate=0 angle=0) order=(0 to 12000 by 1000)
axis2 minor=none label=(angle=90 "aEEG");
symbol1 line=1 ci=black cv=black value=Dot interpol=join width=1 height=1 pt;
title height=14pt 'Asphixiás amplitúdó integrált EEG értékek';
proc gplot data=m767 1 ah2;
      plot asph*time
      / haxis=axis1
        vaxis=axis2
        cframe=white
        caxis=black;
run; quit;
goptions reset=all;
goptions reset=all border ctext=black ftext="Arial" htext=10 pt;
axis1 minor=none label=("Time (0.1 sec)") offset=(5, 5)
      value=(rotate=0 angle=0) order=(0 to 8000 by 1000)
axis2 minor=none label=(angle=90 "aEEG");
symbol1 line=1 ci=black cv=black value=Dot interpol=join width=1 height=1 pt;
title height=14pt 'Amplitúdó integrált EEG értékek az első órában';
proc gplot data=m767 1 ah2;
      plot h1*time
      / haxis=axis1
        vaxis=axis2
        cframe=white
        caxis=black;
run; quit;
goptions reset=all;
```

```
goptions reset=all border ctext=black ftext="Arial" htext=10 pt;
axis1 minor=none label=("Time (0.1 sec)") offset=(5, 5)
value=(rotate=0 angle=0) order=(0 to 7000 by 1000)
axis2 minor=none label=(angle=90 "aEEG");
symbol1 line=1 ci=black cv=black value=Dot interpol=join width=1 height=1 pt;
title height=14pt 'Amplitúdó integrált EEG értékek a második órában';
proc gplot data=m767_1_ah2;
    plot h2*time
    / haxis=axis1
    vaxis=axis2
    cframe=white
    caxis=black;
run; quit;
goptions reset=all;
```

A következő ábrák az előző négy SAS kód lefuttatásával készültek.



## 72 / 9



Asphixiás amplitúdó integrált EEG értékek

## Amplitúdó integrált EEG értékek az első órában





#### Amplitúdó integrált EEG értékek a második órában

## 6 Idő alapú jelfeldolgozás SAS szoftver felhasználásával

#### ARTEFAKTUMOK KISZŰRÉSE 6.1

A láthatóan nagyon kiugró értékek a jelben csak artefaktumok.

Az ilyen zajos adatok kiszűrését a következő rövid kódrészlettel tudjuk megtenni SAS-ban ugyancsak egy Program node-ban, mely során az, hogy a szűrési algoritmusban milyen szorzót alkalmazunk, az egyszerűen állítható (jelen példában 8-as szorzót használunk a második DATA step-ben - szürkével jelölve). Az artefaktum nélküli értékeket a második DATA step-ben létrejött ideiglenes *alap\_no\_artefact* nevű adatállományban az *alap2* nevű változóban tároljuk.

```
/* http://scott.sherrillmix.com/blog/programmer/sas-macros/ alapján */
proc means data=work.m767 1 ah2 mean;
    var alap;
     output out=mean mean=alap mean;
run:
/* mean(alap)=9.39 */
data mean;
```

```
set mean;
call symput('alapmean',trim(left(alap_mean)));
run;
data alap_no_artefact;
set work.m767_1_ah2;
if alap > &alapmean *8 then alap2=&alapmean; else alap2=alap;
run;
```

A következő kódrészlet a már az adatok megjelenítésénél látott GPLOT eljárás alkalmazása az alap mért jelre. Két görbét is megjelenítünk, hogy látszódjon, a szűrés működött. Az eredeti *alap* nevű változót kékkel rajzoljuk ki (SYMBOL1 formája szerint), míg erre rakjuk a szűrt *alap2* változót pirossal (SYMBOL2). A PROC GPLOT-ban a symbol1 és symbol2 utasítások aktiválása a PLOT-ban a per jel előtti számokkal történik (zölddel színezve a háttere a kódban az alap2 beállításának).

Az eredmény egyrészt a számított átlagértéket mutatja, valamint az átlagtól vett 8-szoros különbséggel számolva a kiugró értéket, azokat kiszűrve szemlélteti vonaldiagramon.

Th	e MEANS Procedu	ire
	Analysis Variable : alap	
	Mean	
	9.3905537	



Csak az artefaktum nélküli görbe és SAS kódja:

```
goptions reset=all;
symbol1 interpol=join color=red;
axis1 minor=none label=("Time (0.1 sec)") offset=(5, 5)
       value=(rotate=0 angle=0) order=(0 to 7000 by 1000);
axis2 minor=none label=(angle=90 "aEEG");
title 'Artefaktum nélküli alap aEEG';
proc gplot data=alap_no_artefact;
plot alap2*time=1 / haxis=axis1 vaxis=axis2;
run; quit;
goptions reset=all;
```



## 6.2 EGYENIRÁNYÍTÁS (REKTIFIKÁLÁS)

Egyenirányításra akkor van szükség, ha negatív mért értékeink is lennének. Jelen esetben amplitúdó integrált EEG értékeink vannak, melyekből első lépésben már abszolút érték lett képezve, vagyis megtörtént az egyenirányítás, nincs rá külön szükség.

Ha más jelünk lenne, ahol van értelme rektifikálni, akkor annak a SAS kódja az alábbi lenne. Általános kódot láthatunk, ahol a kiinduló adatállomány *adat* névre hallgat, és ebből egy *adat\_ret* nevű adatállományt hozunk létre, melyben mondjuk a h22 változó jelét rektifikáljuk egy új, r elnevezésű változóba. Az eddigiek alapján ezt is megjeleníthetjük szemléltetés végett egy GPLOT eljárással.

```
/* Rektifikálás: egyenirányítás */
data adat_rekt;
set adat;
if h22<0
    then r=0+(0-h22);
    else r=h22;</pre>
```

run;

Megjegyzés: az egyenirányítás tulajdonképpen egy abszolút érték műveletet jelent matematikailag, ami igencsak adatvesztéssel jár, az adatok szóródása is megváltozik ezzel. Radikális változást eredményez, amit jó szem előtt tartani. Természetesen lehetnek olyan esetek, amikor nincs értelme az irányultságot értékelni csak az amplitúdó nagysága számít, ekkor alkalmazható.

#### 6.3 NORMALIZÁLÁS

Normalizálást követően lesznek a mérések összehasonlíthatóak. Egy adott értékhez (maximum, átlag, tetszőleges) viszonyítunk.

Például a maximum értékhez viszonyítva (ezzel osztva) a 0-1 skálára hozzuk az értékeinket az alábbi SAS kód segítségével:

```
/* Normalizálás:
    maximum értékkel való osztás - rektifikált adatokon. */
proc means data=adat_rekt max;
    var r;
run;
data adat_rekt_norm;
    set adat_rekt;
    normalt = r / 42; /* a max érték 42 */
run;
```

## 6.4 ÁTLAGOLÁSOK

A váltakozójelek jellemzésére gyakran használjuk középértékeiket. Ezek a középértékek átlagértékek. Ezek egyszerű integrálások (adott időegységenként vagy meghatározott feszültséghatárig), illetve középérték számítások (10).

A következőben egy a saját adatainkra aktualizált SAS macro-t olvashatunk, mely elvégzi az integrálszámítást (12).

Ha az előző egyenirányítás és normalizálás lépéseket a 22-edik órában mért első 100 jel adatán végezzük el, és ezen adatok a *h22\_100\_rekt\_norm* nevű adatállományban vannak eltárolva a *normalt* nevű változóban, akkor a következő kódot lefuttatva tudunk integrált számolni ezen görbéhez.

\*http://www.lexjansen.com/wuss/2004/posters/c\_post\_the\_sas\_calculations\_.pdf
\* The following two SQL procedures are optional code blocks, choosing one of
them to compute the baseline value from some early observations of the data file
and store it as a macro variable, BaseY;
\* Baseline value is taken as the mean of all observed values prior to the time
zero.;
PROC SQL;
SELECT MEAN(normalt) FORMAT=6.2 LABEL='Avg normalt' INTO : BaseY

László Anna

```
FROM h22 100 rekt norm
WHERE time LT 0 ;
OUIT
* Baseline value is taken as the mean of the first 3 observed values.;
PROC SQL INOBS=3 ;
SELECT MEAN(normalt) FORMAT=6.2 LABEL='Avg normalt' INTO : BaseY
FROM h22 100 rekt norm;
QUIT;
* This is the macro that calculates the 3 AUCs.;
%MACRO AUC(baseline, dataset, output);
DATA &output;
SET &dataset (WHERE=(time GE 0));
RETAIN Basevalue;
IF & baseline = 0 THEN Basevalue = 0.0;
*&BaseY shown in the following statement is the macro variable defined in any
one of the above SQL procedures;
IF (&baseline = 1 OR &baseline = 2) AND N = 1 THEN Basevalue = &BaseY;
normalt = normalt - Basevalue;
DROP LagTime LagValue;
LagTime = LAG(time);
LagValue = LAG(normalt);
IF time =0
THEN DO;
 LagTime =0;
 LagValue =0;
END:
IF & baseline = 2 AND normalt > 0 AND normalt <= 0.0
THEN DO;
 * Connecting line with positive slope, only the area of right triangle (above
baseline) is counted.;
 DROP Ratio;
 Ratio = normalt / (ABS(LagValue)+normalt);
 Trapezoid = Ratio*(time-LagTime)*(normalt+0.00)/2;
END:
ELSE IF & baseline = 2 AND normalt < 0 AND LagValue >= 0.0
THEN DO;
 * Connecting line with negative slope, only the area of left triangle (above
baseline) is counted.;
 DROP Ratio;
 Ratio = LagValue / (LagValue+ABS(normalt));
 Trapezoid = Ratio*(time-LagTime)*(0.00+LagValue)/2;
END;
ELSE IF & baseline = 2 AND normalt < 0 AND LagValue < 0 THEN Trapezoid = 0.0;
* Negative trapezoidal area is not counted.;
ELSE Trapezoid = (time-LagTime)*(normalt+LagValue)/2 ;
* The rest of all positive trapezoidal areas are counted.;
SumTrapezoid + Trapezoid;
FORMAT Trapezoid SumTrapezoid 8.3 ;
RUN;
%MEND AUC;
```

A SAS Macro meghívása és kiíratása

%AUC(0, h22\_100\_rekt\_norm, integral);
proc print data=integral;

var time normalt sumtrapezoid;
run;

A futtatás eredménye a következő lett:

Obs	time	normalt	SumTrapezoid
1	1	0.38095	0.000
2	2	0.32952	0.355
3	3	0.61905	0.830
4	4	0.49619	1.387
5	5	0.56762	1.919
6	6	0.56762	2.487
7	7	0.44857	2.995
8	8	0.49619	3.467
9	9	0.64286	4.037
10	10	0.33333	4.525
11	11	0.47238	4.928
12	12	0.42857	5.378
13	13	0.47238	5.829
14	14	0.35714	6.243
15	15	0.50000	6.672
16	16	0.44857	7.146
17	17	0.59143	7.666
18	18	0.54381	8.234
19	19	0.47238	8.742
20	20	0.47238	9.214
21	21	0.54762	9.724
22	22	0.80952	10.403
23	23	0.50000	11.058
24	24	0.35714	11.486
25	25	0.61905	11.974
26	26	0.40476	12.486
27	27	0.47238	12.925
28	28	0.42857	13.375
29	29	0.66667	13.923
30	30	0.83333	14.673
31	31	0.42476	15.302
32	32	0.52381	15.776
33	33	0.45238	16.264
34	34	0.32952	16.655
35	35	0.78571	17.213
36	36	0.33333	17.772
37	37	0.37714	18.128
38	38	0.47238	18.552
39	39	0.47238	19.025
40	40	0.40095	19.461
41	41	0.49619	19.910
42	42	0.37714	20.347

43	43	0.45238	20.761
44	44	0.59524	21.285
45	45	0.49619	21.831
46	46	0.49619	22.327
47	47	0.56762	22.859
48	48	0.52000	23.403
49	49	0.47238	23.899
50	50	0.59524	24.433
51	51	0.45238	24.957
52	52	0.35714	25.361
53	53	0.47238	25.776
54	54	0.52000	26.272
55	55	0.52000	26.792
56	56	0.47238	27.289
57	57	0.52000	27.785
58	58	0.52000	28.305
59	59	0.54381	28.837
60	60	0.49619	29.357
61	61	0.35333	29.781
62	62	0.52381	30.220
63	63	0.45238	30.708
64	64	0.52381	31.196
65	65	0.32952	31.623
66	66	0.45238	32.014
67	67	0.52381	32.502
68	68	0.35714	32.942
69	69	0.54762	33.395
70	70	0.32952	33.833
71	71	0.38095	34.189
72	72	0.40095	34.580
73	73	0.49619	35.028
74	74	0.49619	35.524
75	75	0.47238	36.009
76	76	0.38095	36.435
77	77	0.42476	36.838
78	78	0.40095	37.251
79	79	0.49619	37.700
80	80	0.35333	38.124
81	81	0.56762	38.585
82	82	0.50000	39.119
83	83	0.38095	39.559
84	84	0.61905	40.059
85	85	1.00000	40.869
86	86	0.50000	41.619
87	87	0.33333	42.035
88	88	0.47238	42.438
89	89	0.42857	42.889

90	90	0.45238	43.329
91	91	0.47238	43.791
92	92	0.56762	44.311
93	93	0.56762	44.879
94	94	0.45238	45.389
95	95	0.61905	45.925
96	96	0.52000	46.494
97	97	0.32952	46.919
98	98	0.44857	47.308
99	99	0.44857	47.757
100	100	0.47238	48.217

A teljes görbe alatti terület ebben az esetben 48,217 lett.

Megjegyzés: A görbe alatti terület (AUC: Area Under Curve) egyetlen mérőszámmal jellemzi az adott időtartamban vizsgált adatsort. Ez nem írja le a görbe karakterisztikáját. Igen sokféle lefutású görbének lehet ugyanaz az AUC értéke, pedig már ránézésre is jól látható eltérések mutatkoznak. Lehet, hogy már az időintervallum elején magasabb mért értékeink vannak, és ezt követően ez lecseng, vagy pont fordítva, monoton, vagy valamilyen ingadozással, de növekvő trendet mutat. Ezt érdemes szem előtt tartani.

Ez hasonló, mint a modellek összehasonlítására is használatos ROC görbe – Receiving Operating Characteristics -, melynél lehet az AUC értéke ugyanaz pl. 3 modell szenzitivitásának és specifikusságának összevetésekor, mégsem mindegy, hogy melyik görbe milyen sebességgel változik (jellemzően, amelyik inkább közelít a (0,1) ponthoz, az a modell valamivel jobban magyarázza az adatainkat).

## 7 Idősoros elemzés IBM SPSS Statistics felhasználásával

Az alábbi fejezetben az előzőekben SAS-ban mutatott adatfeldolgozást és elemzést SPSS-ben folytatjuk. Az adatbeolvasást, megjelenítést és az artefaktumok kiszűrését természetesen itt is elvégezzük, ezzel összehasonlítva a két szoftverben működő módszerek alkalmazását ezen műveletekre.

A különféle módszerek gyakorlati alkalmazásának szemléltetésére a korábbiakban is használt malac (a kísérletből az egyik hidrogénnel lélegeztetett aszfixiás malac) alap aEEG adatsorát mutatjuk be.

Az idősoros elemzéseket az IBM SPSS Statistics Version 21 szoftverrel végezzük, melyben többnyire a FORECASTING modult futtatjuk. Az analízisek elvégzéséhez elméleti és gyakorlati útmutatót vettünk alapul (13), melyben a 2.10-es fejezetet használtuk fel.

Azt fontos tudnunk, hogy az ilyen sűrű jel feldolgozásában nem feltétlen célravezető az adatok időben történő vizsgálata, ezekkel érdemes viszont kezdeni, majd áttérni különféle spektrális

módszerekkel történő megközelítésekre, de jelen dokumentummal elsősorban az a célom, hogy gyakorlati példákkal szemléltessem az adatelemzési módszerek alkalmazását és értelmezését. A Ketskeméty – Izsó – Könyves Tóth féle könyvben az idősorok elemzéséhez jobban használható napi, heti, havi, negyedéves vagy éves példa adatok állnak rendelkezésre, melyek szabadon elérhetőek. Ilyenek például nyomtatók napi energia-felvételét tartalmazó idősor, részvények jegyzéseinek alakulása, Magyarország népességi adatai egy év havi idősorával, Magyarország villamos energia termelése 1975 és 1994 között vagy például vasúti áruszállítás alakulását mutató negyedéves idősor. Ezekben a cél leginkább valamilyen trend valamint szezonalitás keresése, és a legjobban illeszkedő modell alapján előrejelzés elvégzése. Általános közelítő szabály, hogy egy idősor elemzéséhez minimálisan 50, de inkább néhány száz adatra van szükség (13).

#### 7.1 ADATOK BEOLVASÁSA

Az adatokat ugyancsak a már korábban Excelben előkészített "forSAS" munkalapról olvassuk be, mely az alábbiképpen néz ki MS Excelben:

	🚽 🤊	- 0	-  ∓						r	n76	7_1_	AH2	[ko	mpat	tibilis	üzer	nmód]	- M	icros	oft E	cel						- 0	23
	Fájl	Kezd	őlap	В	eszúi	rás	La	ap el	rend	ezés	e	Ké	plete	k	Adat	ok	Korr	ektúra		Nézet						۵ (	) - 6	23
Be	illesztés Vágólap	∦ ⊡ • ∛	Calib F	ri ⊅ , ✓ ∠ Betű	A → 3> → itípu	• 1   # 	1 1 A A		≡ ≣ ∰ (	≡ ≣ ₽ pazít	≡ ≣ ≫∕ ás		Á S	Italán ,0 ,0 ,0 ,0 ,0 ,0 ,0 ,0 ,0	os % 00 }	• E	Felt Forr Cell	ételes názás astílus Stíli	formá tábláz sok ₹ usok	izás ▼ zatkér	nt •	<b>∃•</b> ≕ B <b>≧*</b> T E E F	eszúrá: örlés 🔻 ormátu Cellák	s * im *	Σ - 	Rendezés és szűrés • Szerkesz	Keresés és kijelölés ♥ tés	
	А	D21		•	0		j	f <sub>x</sub>																				۷
	A	В	С	D	Ε	F	G	Н	Т	J	Κ	L	М	Ν	0	Ρ	Q	R	S	Т	U	V	W	Х	Y	Z	AA	
1	time	alap	asph	h1	h2	h3	h4	h5	h6	h7	h8	h9	h10	h11	h12	h13	h14	h15	h16	h17	h18	h19	h20	h21	h22			
2	1	17	-	91	0	1	9	2	1	3	5	5	7	6	7	24	23	19	6	18	7	3	4	16	16			-
3	2	2 15	16	1	1	2	2	4	2	0	1	5	18	10	0 2	22	10	29	14	11	17	10	16	11	13			-
4		, 4 1 9	- 10	2 1 2 1	1	0	5	- 4	1	8	2	6	7	14	22	10	33	25	7	12	31	9	25	7	20			
6	5	5 17	11	1 1	0	1	1	4	0	6	6	6	9	11	23	2	27	36	26	6	35	13	18	10	3			
7	6	i 8	17	71	0	2	3	6	4	3	15	5	18	7	17	10	6	4	6	15	8	14	50	4	3			
8	7	7 4	11	l 1	1	1	2	0	2	1	7	4	25	3	2	16	6	26	8	19	3	14	21	9	8			
9	8	3 5	6	52	0	1	5	2	2	4	6	15	23	6	6	8	13	25	19	14	8	9	12	30	6			
10	) 9	2	8	31	1	1	1	1	2	2	9	7	35	3	3	4	4	22	25	27	6	4	13	39	27			
11	. 10	) 6	5	5 1	1	1	2	6	2	5	13	8	43	4	6	23	17	44	16	6	8	5	4	31	14			
12	2 11	. 3	1	L 2	1	1	2	1	1	8	5	8	20	6	4	31	24	13	20	11	13	18	11	21	7			-
13	12	2 2	4	1 1	2	1	2	4	1	6	9	9	18	7	6	43	30	5	11	6	7	4	27	32	18			-
14		2	12	2 1	0	2	1	6	4	10	26	6	39	1/	6	45	21	1/	9	1/	14	14	28	8	15			-
10	14	+ 0 : /		2 0	2	1	1	4	10	8 6	10	14	15	2	4	20	22	24	4	34	14	12	29	0 21	21			-
17	16	, ; 8	-	7 1	1	2	1	6	5	2	19	15	16	- 6	21	22	45	10	17	20	12	12	10	14	-21			
18	17	7 3		1 1	0	2	1	4	3	8	10	7	6	8	19	38	38	20	23	28	7	45	6	44	2			
19	18	3 4		1 2	1	0	2	2	2	6	7	6	4	10	20	44	17	22	23	20	14	52	7	53	4			
20	19	7	6	52	1	0	3	1	2	5	16	6	5	13	17	61	9	5	24	18	19	23	2	73	7			
21	. 20	9	2	2 1	1	0	4	1	12	1	2	7	11	9	40	29	7	2	24	31	10	21	9	43	7			•
н	$\bullet \models \models$	nyer	s / M	lunka	a2 ,	Mu	unka3	Z	0-5	6	-10	/1	1-25	/ 26	5-50	/ 51	-100	fo	•									1
Ké	sz																							100	% (=	)		) .:

Jellemzően minden statisztikai, adatelemző szoftver inputjaként hasonló elrendezésre kell az adatokat hozni: a sorokban az egyes megfigyelések, az oszlopokban pedig a változók.

Az első sorba be lehet a változó neveket írni, az adatimport során ezeket jellemzően változó nevekké lehet konvertálni a szoftverekben. A lényeg az, hogy ne használjunk speciális karaktereket, ékezeteket és szóközöket a változók elnevezésekor, mégis beszédesek legyenek a használt nevek. Ezután címkékkel elláthatók minden szoftverben a változók, ahol már fel lehet ékezetes betűket és egyéb karaktereket is használni.

Az adatimporthoz SPSS-ben a következő szintaxist futtathatjuk le, vagy kattintással is beolvashatjuk az adatokat a *File*  $\rightarrow$  *Read Text Data*... menüponttal.



Az SPSS-ben nyitható külön *Syntax* ablak (*File*  $\rightarrow$  *Open*  $\rightarrow$  *Syntax*..., ahol az adott kódrészlet kijelölésével az ikonsoron a nagy zöld háromszöggel (*Run Selection*) lehet a futtatást elvégezni. Itt a *Syntax* ablakban a fenti képen látható módon színezi a kódot a program, hogy könnyebben lehessen értelmezni.

Az importálandó file típusát a TYPE opcióban adhatjuk meg, jelen esetben ez XLS. Az adatokat tartalmazó file fizikai elérési útját a FILE opcióban definiálhatjuk. Azt, hogy a megadott XLS kiterjesztésű file melyik munkalapja tartalmazza az adatállományt, melyet be szeretnénk olvasni, a SHEET opcióban adjuk meg. Mindez a GET DATA utasítás paramétereit képezi. A SAVE utasítással az OUTFILE-ban megadott fizikai elérési útra s név szerint elmenthetjük az SPSS adatállományunkat. Ez jelen esetben SAV kiterjesztést jelent.

A futtatással a következő kép szerint rendben beolvasásra kerültek az adatok az IBM SPSS rendszerbe.

🍓 *m76	7_1_4	H2_01	rigin.:	sav [Da	taSe	t1] - II	BM SP	SS Sta	tistics	Data I	Edito	ľ																×
<u>F</u> ile <u>E</u> di	t <u>V</u> iev	v <u>D</u> at	ta <u>T</u> r	ansform	<u>A</u>	nalyze	Direct	t <u>M</u> arket	ing <u>G</u>	raphs	Utiliti	es	Add-	ons	<u>Wind</u>	low	Help	)										
					2			Ē 🛔			μ	Å	Å	¥,		4					/ 1¢				<b>M</b> 6			
1 : time			1																						1	/isible: 2	5 of 25 Variak	oles
		time	alap	asph	h1	h2	h3	h4	h5	h6	h7	h8	h9	h10	h11	h12	h13	h14	h15	h16	h17	h18	h19	h20	h21	h22	var	
1		1	17	9	1	0	1	9	2	1	3	5	5	- 7	6	- 7	24	23	19	6	18	7	3	4	16	16		
2		2	15	7	1	1	2	2	4	2	6	1	8	18	10	6	22	16	29	8	3	22	10	5	11	13		
3		3	4	16	1	1	0	5	4	1	9	2	5	19	21	3	- 7	37	49	14	11	17	11	16	7	26		
4		4	9	3	1	1	0	5	5	1	8	3	6	- 7	14	22	10	33	25	7	12	31	9	25	7	6		
5		5	17	11	1	0	1	1	4	0	6	6	6	9	11	23	2	27	36	26	6	35	13	18	10	3		
6		6	8	17	1	0	2	3	6	4	3	15	5	18	7	17	10	6	4	6	15	8	14	50	4	3		
7		7	4	11	1	1	1	2	0	2	1	7	4	25	3	2	16	6	26	8	19	3	14	21	9	8		
8		8	5	6	2	0	1	5	2	2	4	6	15	23	6	6	8	13	25	19	14	8	9	12	30	6		
9		9	2	8	1	1	1	1	1	2	2	9	7	35	3	3	4	4	22	25	27	6	4	13	39	27		
10		10	6	5	1	1	1	2	6	2	5	13	8	43	4	6	23	17	44	16	6	8	5	4	31	14		
11																												
12		12	2	4	1	2	1	2	4	1	6	9	9	18	7	6	43	30	5	11	6	7	4	27	32	18		
13		13	2	12	1	0	2	1	6	4	10	26	6	39	17	6	45	21	17	9	17	14	14	28	8	7		
14		14	6	9	1	2	1	1	4	10	8	16	14	15	5	4	20	22	24	4	34	14	2	29	8	15		
15		15	4	8	0	0	1	0	3	8	6	9	11	11	4	15	69	36	5	11	6	12	12	21	21	21		
16		16	8	7	1	1	2	1	6	5	2	19	15	16	6	21	22	45	10	17	20	12	12	10	14	8		
17		17	3	4	1	0	2	1	4	3	8	10	7	6	8	19	38	38	20	23	28	7	45	6	44	2		-
	1																	-				_		_		_		
Data Vie	w Va	riable V	iew																									
																					BM SPS	S Statis	tics Pro	cessor i	is ready			

Ellenőrizzük, hogy minden változót érték jellegű adatként ismert-e föl az SPSS. A *Variable View* munkalapon minden változó (itt sorokban találjuk) legyen skálaváltozó, vagyis a *Measure* oszlopban legyen *Scale* beállítva a következő kép szerint.

🍓 *m767_	1_AH2_origin.s	av [DataSet1]	BM SPSS	Statistics Da	ta Editor									
<u>F</u> ile <u>E</u> dit	<u>⊻</u> iew <u>D</u> ata <u>T</u> ra	ansform <u>A</u> nalyz	e Direct <u>M</u> a	rketing <u>G</u> rap	hs <u>U</u> tilities Add-g	ons <u>Wi</u> ndow	Help							
🔁 H	🖨 🛄		¥ 🎬		H M		- S		4 📀 🌑	ABG				
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role			
1	time	Numeric	11	0		None	None	4	🗮 Right	🔗 Scale	ゝ Input 🖆			
2	alap	Numeric	11	0		None	None	3	🗮 Right	🔗 Scale	ゝ Input			
3	asph	Numeric	11	0		None	None	3	🗮 Right	🗸 Scale	🔪 Input			
4	h1	Numeric	11	0		None	None	2	🗮 Right	🔗 Scale	🕥 niput			
5	h2	Numeric	11	0		None	None	3	🗮 Right	🛷 Scale 🛛 🔻	ゝ Input			
6	h3	Numeric	11	0		None	None	3	🗮 Right	🔗 Scale	ゝ Indut			
7	h4	Numeric	11	0		None	None	3	🗮 Right	🔗 Scale	Input			
8	h5	Numeric	11	0		None	None	3	🗮 Right	Coale	💊 Input			
9	h6	Numeric	11	0		None	None	3	🚎 Right	🔗 Scale	ゝ Input			
10	h7	Numeric	11	0		None	None	2	🗮 Right	🔗 Scale	ゝ Input			
11	h8	Numeric	11	0		None	None	2	🗮 Right	🔗 Scale	ゝ Input			
12	h9	Numeric	11	0		None	None	2	🗮 Right	🔗 Scale	ゝ Input			
13	h10	Numeric	11	0		None	None	2	🚎 Right	🔗 Scale	ゝ Input			
14	h11	Numeric	11	0		None	None	2	🗮 Right	🔗 Scale	ゝ Input			
15	h12	Numeric	11	0		None	None	2	🚎 Right	🔗 Scale	ゝ Input			
16	h13	Numeric	11	0		None	None	2	🗮 Right	🔗 Scale	ゝ Input			
17	h14	Numeric	11	0		None	None	2	🚎 Right	🛷 Scale	ゝ Input			
18	h15	Numeric	11	0		None	None	3	🗮 Right	🔗 Scale	ゝ Input			
19	h16	Numeric	11	0		None	None	3	🗮 Right	🔗 Scale	ゝ Input			
20	h17	Numeric	11	0		None	None	3	🗮 Right	🔗 Scale	ゝ Input			
21	h18	Numeric	11	0		None	None	3	🗮 Right	🔗 Scale	ゝ Input			
22	h19	Numeric	11	0		None	None	3	🗮 Right	🔗 Scale	ゝ Input			
23	h20	Numeric	11	0		None	None	3	🗮 Right	🔗 Scale	ゝ Input			
24	h21	Numeric	11	0		None	None	3	🗮 Right	🔗 Scale	ゝ Input			
25	h22	Numeric	11	0		None	None	3	🚎 Right	🛷 Scale	ゝ Input			
26	26													
	1													
Data View	Variable View	)												
								uma / ==						

## 7.2 ADATOK SZŰRÉSE

Ahogy azt a fejezet elején kifejtettük, idősoros elemzésben jellemzően legalább 50, de inkább néhány 100 adat szükséges az adatelemzéshez. A mért aEEG jel nekünk nagyobb frekvencián mért, az alap jelből kb. 6000 adatunk van. A módszerek és azok értelmezésének ismertetése és könnyebb áttekinthetőség végett csak másodperc alapon tartjuk meg az adatokat. Ezért kiszűrjük minden tízedik adatot, hiszen tized másodpercenként van mért jelünk az adatállományban. Ehhez a következő kódsort futtathatjuk le, mellyel egyúttal a kiszelektált adatokat új adatállományba is mentjük, és az előzőt be is zárhatjuk.



Mivel kb. minden óra első 10 percében történt a jel detektálása és mentése, de ez nem pontosan 6000 adatot jelent, most levágjuk a 10 perc fölötti esetleges adatokat, a módszerek futtatása és értelmezése szempontjából nem számítanak. Ezt a fenti szintaxisból a következő sor végzi:

SELECT IF (MOD(time,10)=0 and time<=6000).

A Data View nézetben az SPSS-ben a következő kép szerint sikeresen leválogattuk az adatainkat.

Visible: 25 of 25 V													25 of 25 Varial	ples													
	time	alap	asph	h1	h2	h3	h4	h5	h6	h7	h8	h9	h10	h11	h12	h13	h14	h15	h16	h17	h18	h19	h20	h21	h22	var	
588	5880	7	3	2	1	1	0	0	4	4	5	4	5	50	51	41	9	43	43	43	38	15	12	23	6		
589	5890	32	1	0	1	1	1	2	7	2	2	8	23	11	- 7	4	8	14	13	5	26	7	17	13	14		
590	5900	- 7	1	1	0	1	1	11	5	3	3	22	10	24	46	8	11	23	4	24	14	32	8	35	19		
591	5910	12	10	1	0	1	2	2	9	6	6	9	1	4	9	4	12	34	5	7	13	40	28	12	7		
592	5920	8	4	1	1	1	2	1	2	5	8	5	5	12	9	10	5	29	21	14	10	23	23	17	18		
593	5930	9	5	0	1	1	3	1	1	- 7	11	31	42	13	46	20	15	3	11	16	39	13	24	12	3		
594	5940	5	4	1	0	2	0	6	3	13	26	2	2	6	20	5	30	10	11	34	24	33	6	13	20		
595	5950	25	3	1	0	1	1	1	1	3	28	4	17	6	26	15	7	36	21	24	35	26	12	14	10		
596	5960	4	12	0	1	1	2	1	7	8	19	16	12	1	2	47	13	6	31	12		57	17	5	7		
597	5970	12	6	5	1	2	24	1	4	30		9	6	14	3	34	52	15	6	6		15	12	27	11		
598	5980	6	8	0	0	0	2	2	7	1		- 7	11	5	13	25	7	13	19	5		11	8	11	16		
599	5990	10	2	2	0	1	1	13	9	3		3	- 7	31	7	7	18	71	23	1		12	43	19	23		
600	6000	6	2	2	0	0	2	1	3	9		9	43	4	15	16	4	21	62	6		36	19	5	20		
601																											

Jól látható, hogy mind a 25 változóban (oszlopok) 600 másodpercre van adatunk (sorok), mindössze két órában történt a jel regisztrálása rövidebb, mint 10 percig (h8 és h18 – ezekben hiányzó értékeket látunk az utolsó sorokban, de ennyi adat mellett a 4-5 hiányzó érték elenyésző).

A *Variable View* nézetben most is leellenőrizhetjük, hogy minden változó maradjon meg mérési skálán értelmezettnek, de itt már, mivel előzőleg a teljes adatállományon beállítottuk, nincs rá szükség külön.

A fenti adatszűrés természetesen menüből is vezérelhető kattintással, ehhez a *Data* menüben a *Select Cases*... menüpontot kell választani.

#### 7.3 MEGJELENÍTÉS

Az alap jel másodperces adatainak megjelenítéséért a következő programkód felel.

\*\*\*\*\*\*\*Megjelenites. \* Sequence Charts. **TSPLOT VARIABLES=alap** /ID=time /NOLOG /FORMAT NOFILL NOREFERENCE.

Ezt amúgy az Analyze  $\rightarrow$  Forecasting  $\rightarrow$  Sequence Charts... menüponttal is beállíthatjuk.

A TSPLOT utasításban a VARIABLES paraméterben adhatjuk meg a kirajzolandó idősorokat. Most csak egyet, az alap jelet adtuk meg, így azt teszi a függőleges tengelyre a vízszintes ID opcióban megadott time szerint.

A futtatás eredményeként az Output ablakban megkaptuk a következő grafikont.



Jól látható, hogy a 141-edik másodperc környékén van egy kiugró érték (ahogy azt már korábban a SAS szoftverrel is kimutattuk). Ez artefaktumnak minősül, amit ki kell szűrni a jelből. A következő lépésben ezt tesszük meg.

## 7.4 ARTEFAKTUMOK KISZŰRÉSE

SPSS-ben lehetőségünk van anomáliák automatikus detektálására. Erre szolgál a következő kódrészlet, mellyel 3 csoportot képez a szoftver az alap jel értékei alapján, és lementi 1-es, 2-es és 3-as értékekkel egy új, **Peer\_alapId** nevű változóba (ez elérhető a *Data*  $\rightarrow$  *Identify Unusual Cases*... menüponttal is).

	********zaj kiszurese.
	*Kiugro ertekek, anomaliak detektalasa.
	*peer csoportokat kepez. A 3-asok helyett azokat 5-tel megelozo ertek bemasolasa.
	* Identify Unusual Cases.
P	DETECTANOMALY
	/VARIABLES SCALE=alap ID=time
	/PRINT ANOMALYLIST CPS
	/SAVE PEERID(Peer_alapId)
	/HANDLEMISSING APPLY=NO
	/CRITERIA PCTANOMALOUSCASES=5 ANOMALYCUTPOINT=2 MINNUMPEERS=1 MAXNUMPEERS=15 NUMREASONS=1.

A fenti szintaxis futtatásával (a SAVE hatására) létrejön az új változó. Láthatjuk az adatállományban, hogy a nagy kiugró értékünk a 141-edik sorban található, ahogy azt előzőleg a szekvencia diagramon detektáltuk.

141 : alap		492 Visible: 26 of 26 V															ariables											
	time	alap	asph	h1	h2	h3	h4	h5	h6	h7	h8	h9	h10	h11	h12	h13	h14	h15	h16	h17	h18	h19	h20	h21	h22	Peer_alapId	var	
136	1360	5	9	0	6	1	13	1	16	16	- 7	1	12	8	3	- 7	3	11	28	5	26	9	47	4	27	2		1
137	1370	5	1	11	2	18	1	0	20	4	10	4	8	39	6	15	33	23	13	9	6	8	14	32	13	2		
138	1380	12	10	0	2	1	2	2	3	1	1	3	4	3	13	12	30	19	28	23	19	18	11	5	23	1		
139	1390	8	3	0	1	0	1	1	9	5	20	8	5	10	11	24	8	28	10	7	11	17	7	19	16	1		
140	1400	10	1	2	1	2	1	4	2	3	5	29	10	10	1	12	10	17	55	60	8	7	22	8	28	1		
141	1410	492	4	0	1	2	11	5	2	8	14	2	6	19	75	3	12	47	19	33	7	15	13	5	9	3		
142	1420	53	2	1	1	1	2	1	1	27	8	26	8	10	10	30	41	44	23	14	47	14	19	14	9	3		
143	1430	2	1	1	1	0	1	3	4	1	24	8	4	5	9	20	11	22	35	27	4	6	8	32	12	2		
144	1440	8	1	2	1	0	2	3	3	6	10	4	20	4	- 7	9	11	12	18	26	22	5	8	14	18	1		
145	1450	5	1	0	4	2	2	1	3	14	25	23	22	15	11	36	10	8	9	5	25	2	27	38	7	2		
146	1460	3	1	2	5	1	2	0	1	6	6	8	11	9	27	5	24	21	18	6	8	12	11	17	31	2		
147	1470	3	2	3	2	0	1	14	1	29	19	26	26	20	24	- 7	13	20	23	11	6	34	5	14	37	2		
148	1480	12	2	2	2	0	2	2	14	8	4	8	4	3	21	35	16	7	4	14	13	13	21	5	3	1		
149	1490	5	2	1	7	0	1	1	3	13	12	5	3	4	14	10	6	11	32	33	6	67	27	19	8	2		
150	1500	3	3	1	15	2	2	4	8	2	5	34	31	13	27	6	8	21	4	15	30	12	18	5	12	2		
151	1510	11	2	1	2	1	1	1	6	3	5	3	13	24	7	37	5	12	9	6	34	10	31	14	7	1		Ţ
	4												1.1.55															
Data View	Variable V	/iew																										

Kérjünk le alapstatisztikát az új csoport változóról, hogy egy kicsit megismerjük.

 \*anomaliak alapstatisztikai.
 EXAMINE VARIABLES=alap BY Peer\_alapId /ID=time
 /PLOT BOXPLOT STEMLEAF
 /COMPARE GROUPS
 /STATISTICS DESCRIPTIVES
 /CINTERVAL 95
 /MISSING LISTWISE
 /NOTOTAL. A kapott eredményben leolvashatjuk a *Case Processing Summary* táblából, hogy a 3 csoportban hogyan oszlanak el az adatok. Az 1-esbe 341 másodperc adatát sorolta a rendszer, a 2-esbe 247 dbot, míg a 3-asba mindössze 12-t. A fenti *Data View*-ban is látható, hogy a 141-edik másodperc nagyon kiugró 492-es értékét a 3-as csoportba sorolja. Ezt a *Descriptives* táblából is kiolvashatjuk az eredmények közül. Az 1-es csoport 7-27 közötti aEEG alap adatokat tartalmaz, a 2-es a 0-6 intervallumból származóakat, míg a 3-as 29 és 492 közöttieket.

Az átlag standard hibája az első két csoportban 0,5 alatti, vagyis kicsit szóródnak az adatok az átlag körül, míg a 3-as csoportban 38,132, jelezve, hogy itt elég nagy a variancia.

A boxplotról is látható, hogy a 3-as csoportban egy igazán kiugró érték van, a korábban már észlelt 492-es 141 másodpercnél.



Ezért csak a 3-as csoportba tartozó értékeket szűrjük ki az alábbi algoritmus szerint.



Az ábrák alapján láthattuk, hogy a kiugró értékek melletti környező adatok már nem lógnak ki számottevően, így az adott, 3-as csoportba tartozó értékek esetén, az őket 5 másodperccel megelőző értékkel helyettesítjük a kiugró értékeket. Megtehetnénk azt is, hogy látva a görbét, csak a 141-es sor adatát cseréljük le, de egyöntetűen is végezhetünk helyettesítést, hisz a többi adatba ez nem nagyon zavar be, lévén a környező adatok eléggé hasonló amplitúdóval bírnak. Az artefaktum nélküli amplitúdó értékeket egy új, *alap\_c* elnevezésű változóba mentjük, ami természetesen a legtöbb esetben (1-es és 2-es csoportok a Peer\_alapId szerint) az *alap* jellel egyenlő.

Ezeknél a másodperces adatoknál a malac aEEG jelében nem várunk időbeli eltolódást különösképp, azért megnézzük mit mutat az elemzés (alátámasztja-e ezt a megérzésünk).

A fenti szintaxis szerint már meg is jelenítettük a TSPLOT utasítással az eredményt.

Az első diagram az eredeti (kék görbe) és az artefaktum kiszűrésével keletkezett idősort (zöld görbe) is tartalmazza, míg a második már csak a javított tartományra koncentrál.



Láthatjuk, hogy a nagy zaj kiszűrése megtörtént az *alap* jelből *alap\_c* változóba mentve.



Az így kapott szűrt és zajmentes jellel, mint idősorral dolgozunk tovább.

Természetesen ilyen esetben célszerű a szakemberrel egyeztetni, aki a kísérletet tervezte és/vagy végezte, felhasználja az eredményeit, hogy milyen szinten számít artefaktumnak egy-egy adat, és a tisztított idősor már tekinthető-e artefaktumoktól mentes, további elemzésre alkalmas regisztrátumnak.

#### 7.5 NORMALITÁS ELLENŐRZÉS

Érdemes ellenőrizni, hogy az előzőleg korrigált alap jel (*alap\_c* változó értékei) normális eloszlást követ-e, hogy megnézzük, lehet-e benne valamiféle trend, vagy egymástól független adatokról van szó. A következő lépésben tehát megvizsgáljuk, hogy egyáltalán lehet-e az *alap\_c* idősorában trend, vagy ez az idősor időben nem változik. Ehhez azt kell megnéznünk, hogy normális eloszlást követ-e. Ha igen, akkor nem érzékelhető benne időbeni tendencia.

A következő szintaxis lefuttatásával egy P-P diagramot rajzolhatunk ki.

```
*P-P plot.

PPLOT

/VARIABLES=alap_c

/NOLOG

/NOSTANDARDIZE

/TYPE=P-P

/FRACTION=BLOM

/TIES=MEAN

/DIST=NORMAL.
```

A futtatás eredménye az alábbi diagram.



Látható, hogy a pontok nem igazán illeszkednek az átló egyenesére, hanem attól eltérően, egy másik függvény vonulatát követik. Van tehát valamiféle szabályosság az adatsorunkban. Eszerint az a sejtésünk, hogy nem követ normális eloszlást az adatsor.

De nézzük meg egy hisztogramon is az eloszlás alakulását a következő kódrészlet segítségével.



Az eredményt az Output ablakban találjuk.



A hisztogramon is látható a normális eloszlás haranggörbéjétől való eltérés: jobbra ferde eloszlást követnek az értékeink.

Egy statisztikai próbával is alátámaszthatjuk valamelyest a megérzésünket. Futtassunk egy Kolmogorov-Smirnov próbát a normális eloszlástól való eltérés tesztelésére a következő programkód alapján.

\*Kolmogorov\_Smirnov proba. \*Nonparametric Tests: One Sample. **NPTESTS** /ONESAMPLE TEST (alap\_c) KOLMOGOROV\_SMIRNOV(NORMAL=SAMPLE) /MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE /CRITERIA ALPHA=0.05 CILEVEL=95.

A kapott p-érték 0,000 (természetesen nem pontosan 0, csak 3 tizedes jegy pontossággal adja meg a program alapértelmezetten); ami azt jelenti, hogy szignifikánsan eltérünk attól a nullhipotézistől, miszerint normális eloszlást követnek az adataink, vagyis nem normális eloszlású populációból származik ez az idősor.

Azt azért szem előtt kell tartani, hogy ez a próba alapvetően független adatokat feltételez, és a mi esetünkben az idő, mint kovariáns mindig jelen van, eszerint összetartozóak az adataink. De a grafikonok alapján mondhatjuk, hogy van értelme idősoros elemzést végezni az adatokon.

## 7.6 AUTOKORRELÁCIÓ

Egy idősorban az adatok időbeli összefüggését autokorrelációval lehet vizsgálni. Az autokorrelációs és parciális autokorrelációs függvények (**ACF**: AutoCorrelation Function, **PACF**: Partial AutoCorrelation Function) értékeivel lehet felmérni, hogy az idősor adott elemére milyen mélységben van hatással a múltja, azaz a korábbi időpontbeli változók közül mennyi van szignifikáns korrelációban vele (*13*).

Ezek az értékek is, mint a korrelációs koefficiens -1 és 1 közötti értékek lehetnek. A 0-hoz közeliek a lineáris kapcsolat hiányát mutatják, míg az abszolút értékben 1-hez közeliek erős függvényszerű kapcsolatot jelentenek.

SPSS-ben az ACF függvénnyel lehet autokorreláció elemzést végezni. Ez elérhető az *Analyze*  $\rightarrow$  *Forecasting*  $\rightarrow$  *Autocorrelations*... menüponttal is, de a következő szintaxissal (ACF utasítás) is futtatható.

\*\*\*\*\*Autokorrelacio elemzes: az idoben valo osszefugges, kapcsolatok detektalasara.
\*LAG number = 16. (16-os eltolassal).
ACF VARIABLES=alap\_c
/NOLOG
/MXAUTO 16
/SERROR=IND
/PACF.
\*LAG number = 599.
ACF VARIABLES=alap\_c
/NOLOG
/MXAUTO 599
/SERROR=IND
/PACF.

Szokásosan a VARIABLES opcióban adjuk meg az idősort, amire szeretnénk autokorrelációt számítani. Itt a LAG érték az eltolást jelenti, alapértelmezetten 16-ra van beállítva. Az MXAUTO opcióval módosíthatjuk ennek az értékét. Mivel 600 értékünk van alap\_c változóban, 599 eltolás lehet ezen 600 érték között. Az így kapott ACF és PACF grafikonok, vagy más néven **korrelogram**ok már kevésbé áttekinthetőek a túl sok érték miatt. Ugyanakkor érdemes első körben a teljes idősorra megnézni ezeket a grafikonokat egy-egy elemzés során, hogy legyen egy impressziónk az idősor összefüggésének karakterisztikájáról (megjegyzés: LAG=600-zal is lefut). Utána már jellemzően csak olyan LAG értékkel jelenítik meg a korrelogramot, ahányad rendű ARMA (vagy ARIMA) modellt illesztenek az adatokra (ez jellemzően maximum 3).

A korrelogramok vizsgálata ad egy elsődleges benyomást, hogy milyen ARIMA modellt érdemes alkalmaznunk majd az idősorunk sztochasztikus kapcsolatainak feltárására. A modellillesztés felismerésében támogatnak ezek.



Azért már ezen ábrákról is leolvasható az, hogy nincs igazán olyan érték, ami nagyon túllógna a konfidencia intervallumot szemléltető vonalakon (konfidencia sáv), jelezve, hogy annyi másodperc eltolással szignifikáns a kapcsolat az időben.

Egy 100-as LAG érték beállításával is megjeleníthetjük az ACF és PACF függvényeket egy jobb fókusz érdekében.





A parciális autokorrelációs értékek néhol lépik túl a megadott görbét, de csak igen kicsi értékekkel. Az autokorrelációs értékek -0,084 es 0,11 között mozognak, vagyis igen kicsik. Eszerint nem mutatható ki erős (értékelhető) autokorreláció (időben történő összefüggés) az alapsorban (korrigált értékében:  $alap_c$ ).

Alapvetően a statisztikai elemzések értelmezésekor mindig érdemes figyelni arra, hogy milyen szakmai kérdést is kívánunk megválaszolni éppen a statisztika eszközrendszerével. Attól, hogy valami szignifikáns eredményt mutat, még lehet, hogy szakmailag nem releváns a kimutatott eltérés (hatás, stb.). Figyeljünk mindig, hogy ne értelmezzük félre ezeket az eredményeket. Az is igaz mindig, hogy az elemszám növelésével ugyanazon eltérés már egy idő után szignifikáns lesz, de, ha szakmailag nem értelmezhető, akkor ne vegyük figyelembe.

## 7.7 GÖRBEILLESZTÉS

Azt már a korábbi idősor görbéken is láttuk, hogy az artefaktumoktól mentes korrigált alap jelben sűrűn váltják egymást a csúcsok és a völgyek. Talán valamiféle nagyon lassú emelkedés észlelhető az értékekben időben. Nézzük meg első körben egy lineáris görbe illesztését a pontjainkra.

A következő kódrészlet tulajdonképpen egy lineáris trend és egy hibatag összegéből álló dekompozíciós modellt illeszt az értékeinkre.

\*\*\*\*\*\*\*Egyszeru gorbeillesztes.
\*Dekompozicios modell: alap aEEG elorejelzese egy linearis trend es egy hibatag osszegevel.
\* Curve Estimation.
TSET MXNEWVAR=1.
PREDICT THRU end.
CURVEFIT
/VARIABLES=alap\_c
/MODEL=LINEAR
/PRINT ANOVA
/PLOT FIT
/SAVE=PRED.

Az eredményből kiolvashatók az alábbiak:

A *Model Summary* tábla alapján az illesztés nem sikeres, mert az R-négyzet értéke mindössze 0.007 (a lineáris trend szinte semmit nem magyaráz meg a jelenségből, hiszen az adatok varianciáját csupán 0,7%-ban magyarázza). A legtöbbször használják az R-négyzet értéket modell magyarázó erejének meghatározásakor, ugyanakkor itt is, mint a nemparaméteres Kolmogorov-Smirnov próbánál figyelembe kell vennünk, hogy független adatok esetén működik, ami idősor esetén nem áll fenn. Jellemzőbb az AIC (Akaike Information Criterion), vagy a BIC (Bayes Information Criterion) értékek alapján való döntés (több modell futtatásakor a kisebb AIC vagy BIC értékkel rendelkező modellt érdemes választani), ezek ugyanakkor még nincsenek minden hasonló próbához beépítve a statisztikai programokba. SPSS sem kezeli ezt a CURVEFIT utasításában.

A *Coefficients* táblában találjuk, hogy az illesztett egyenes meredeksége/iránytangense: 7,774. Ezzel a hozzá tartozó t tesztstatisztika p-értéke (p=0.000) alapján a meredekség szignifikánsan különbözik a 0-tol, de igen kicsi ez az érték, nem nagyon értelmezhető.



Az illesztett egyenesen látható az előzőleg a számokkal is alátámasztott nagyon enyhe emelkedés az időben.
A futtatás eredményeként kimentettük az illesztett lineáris modell értékeit, melyet az adatállományban egy *FIT\_1* értékkel tárol az SPSS. Ezt a *Variable View* nézetben átírhatjuk mondjuk *fit\_linear\_alap*-ra a következő ábrák szerint (előbb *Data View*-ban az értékek még FIT\_1 néven, majd *Variable View*-ban már az átírt változónév).

<b>h *</b> m767	_AH2	2_origi	n_time	_sec.	sav [	m76	7_AI	12_0	rigin	] - 11	BM SP	PSS St	atist	ics Da	ta Edi	tor									
<u>F</u> ile <u>E</u> dit	⊻iev	∾ <u>D</u> ata	a <u>T</u> ran	nsform	<u>A</u> n	alyze	Dir	ect <u>M</u>	arketin	g g	<u>G</u> raphs	s <u>U</u> ti	lities	Add-g	ins V	<u>V</u> indow	Help								
🔁 l												μ			*,	4		4	2		A 14	•	ABG		
1 : FIT_1			7,776	489184	69218	в																	V	/isible: 28 of 28	3 Variables
		h4	h5	h6	h7	h8	h9	h10	h11	h12	h13	h14	h15	h16	h17	h18	h19	h20	h21	h22	Peer alapId	alar	) C	FIT 1	va
1	1	2	6	2	5	13	8	43	4	6	23	17	44	16	6	8	5	4	31	14	2		6,00	7,77649	-
2	0	4	1	12	1	2	7	11	9	40	29	7	2	24	31	10	21	9	43	7	1		9,00	7,77922	
3	2	1	3	6	2	7	40	8	11	3	39	27	9	7	11	18	51	17	3	35	1		7,00	7,78196	
4	1	23	1	6	2	7	10	21	11	5	33	12	37	3	4	10	18	8	22	10	2		4,00	7,78469	
5	1	3	2	26	28	1	4	21	4	20	17	4	13	42	5	10	14	16	7	25	1		8,00	7,78742	
6	2	1	39	1	6	4	3	7	5	5	6	11	40	5	5	17	22	16	11	6	2		4,00	7,79015	
7	1	2	10	2	5	10	7	32	35	8	5	7	12	3	26	12	46	9	2	13	2		3,00	7,79289	
8	1	1	3	2	14	7	19	63	20	39	5	12	16	12	14	12	10	11	14	12	2		4,00	7,79562	
9	2	2	7	2	4	19	5	5	8	12	5	17	18	15	21	8	26	16	3	19	1		7.00	7,79835	
10	0	2	2	2	10	17	4	11	5	2	38	31	14	7	12	13	6	16	13	7	1		10.00	7.80109	
11	0	2	4	4	4	5	14	5	6	11	7	11	44	18	4	6	32	10	16	40	2		6.00	7.80382	
12	1	2	2	22	30	15	5	5	19	8	7	12	14	17	10	7	4	10	26	6	2		4.00	7.80655	
13	1	3	1	1	3	6	5	2	16	6	12	9	3	10	36	. 7	2	12	13	17	2		6.00	7 80928	
	Data View Variable View																								
Data View	Data View Variable View																								
	IBM SPSS Statistics Processor is ready																								
ta m767	*m767_AH2_origin_time_sec.sav [m767_AH2_origin] - IBM SPSS Statistics Data Editor																								
File Edit	le Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help																								
j 🔁 🖥	Image: Image																								
	Name         Type         Width         Decimals         Label         Values         Missing         Columns         Align         Measure         Role																								
17	h1-	4		Num	eric	11	0												None	None	2	Right	Scale	🔪 Input	<u></u>
18	h1:	5		Num	eric	11	0												None	None	3	Right	Scale	> Input	- 1
19	h10	0 7		Num	eric oric	11	0												None	None	3	Eight	Scale	> Input	- 1
20	h1	r B		Num	eric	11	0												None	None	3	Right	Scale	> Input	-
22	h19	9		Num	eric	11	0												None	None	3	🚍 Right	Scale	Input	
23	h2(	- D		Num	eric	11	0												None	None	3	🔳 Right	Scale	> Input	
24	h2	1		Num	eric	11	0												None	None	3	🔳 Right	Scale	🖒 Input	
25	h2:	2		Num	eric	11	0												None	None	3	🗃 Right	🔗 Scale	🔪 Input	
26	Pe	er_alapl	d	Num	eric	8	0		F	eer	Group	ID							None	None	8	🔳 Right	🚴 Nomin	al 🖒 Input	
27	ala	ip_c		Num	eric	8	2												None	None	10	🗃 Right	🛷 Scale	ゝ Input	
28	fit_	linear_a	lap	Num	eric	11	5		F	it for	alap_	c from	CUR	RVEFIT,	MOD_	8 LINEA	R		None	None	7	🗐 Right	🧳 Scale	🦒 Input	
29																									
20	1																						1		4
Data View																									
Data Hoff	Varia	able Viev	N																						
	Varia	able Viev	N																		IBM CDCC C4	stistice Dros			

A *Label* oszlopban bármely címke adható a változóknak. Az SPSS a modellillesztés során a fenti képen látható címkét generálja.

Nézzünk meg egy négyzetes és egy köbös polinom illesztést is, bár az idősort látva nem várunk sokkal jobb illeszkedést. Ehhez az alábbi kódot futtassuk le a *Syntax* ablakban.



Láthatjuk, hogy szinte ugyanazt a programsort kell futtatni, csak a MODEL opcióban adunk meg más paramétert a négyzetes (QUADRATIC) és köbös (CUBIC) polinom illesztéshez.

Két új változót képzünk: a két illesztett függvény szerinti értékeket. Ezeket az előbbi, lineáris modell mintájára hasonlóan átnevezhetjük, ahogy azt a következő kép is mutatja.

m767_A	m767_AH2_origin_time_sec.sav [m767_AH2_origin] - IBM SPSS Statistics Data Editor														
<u>File Edit (</u>	⊻iew <u>D</u> ata <u>T</u> ransfo	m <u>A</u> nalyze	Direct	t <u>M</u> arketing	<u>G</u> raphs <u>U</u> tilities	Add- <u>o</u> r	ns <u>Wi</u> ndow	Help							
		5 7	1	i 📥	判	ÅÅ		- S				-	ò		
	Name	Туре	Width	Decimals			Label			Values	Missing	Columns	Align	Measure	Role
17	h14	Numeric	11	0						None	None	2	)를 Right	🔗 Scale	ゝ Input
18	h15	Numeric	11	0						None	None	3	🗮 Right	🛷 Scale	ゝ Input
19	h16	Numeric	11	0						None	None	3	🗮 Right	🔗 Scale	🔪 Input
20	h17	Numeric	11	0						None	None	3	🔳 Right	🔗 Scale	🔪 Input
21	h18	Numeric	11	0						None	None	3	🔳 Right	🛷 Scale	🔪 Input
22	h19	Numeric	11	0						None	None	3	🔳 Right	🛷 Scale	🔪 Input
23	h20	Numeric	11	0						None	None	3	🔳 Right	🛷 Scale	🔪 Input
24	h21	Numeric	11	0						None	None	3	🔳 Right	🛷 Scale	🔪 Input
25	h22	Numeric	11	0						None	None	3	🔳 Right	🛷 Scale	🔪 Input
26	Peer_alapId	Numeric	8	0	Peer Group ID	)				None	None	8	🔳 Right	\delta Nominal	🔪 Input
27	alap_c	Numeric	8	2						None	None	10	🔳 Right	🛷 Scale	🔪 Input
28	fit_linear_alap	Numeric	11	5	Fit for alap_c	from CUI	RVEFIT, MOD	8 LINEAR		None	None	7	🔳 Right	🛷 Scale	🔪 Input
29	fit_quad_alap	Numeric	11	5	Fit for alap_c	from CUI	RVEFIT, MOD	9 QUADRATI	С	None	None	13	🔳 Right	🛷 Scale	🔪 Input
30	fit_cubic_alap	Numeric	11	5	Fit for alap_c	from CUI	RVEFIT, MOD	9 CUBIC		None	None	13	🔳 Right	🛷 Scale	🔪 Input
31															

Az eredményből kiolvashatjuk az R-négyzet értékeket a modell magyarázó erejéről (ismét tartsuk szem előtt, hogy ez független adatok esetén értelmezhető!). Mindkét érték igen kicsi (négyzetes esetén 0,007, köbös esetén 0,014).

Ha szemléltetjük az előző 3 modellillesztést a már módosított változóneveket alkalmazzuk a TSPLOT függvény VARIABLES utasításában.

\*szemleltetes. **TSPLOT VARIABLES**=alap\_c fit\_linear\_alap fit\_quad\_alap fit\_cubic\_alap /ID=time /NOLOG.

Az eredmény az alábbi.



A 3 illesztett görbe szinte egybeesik, és egyik sem illeszkedik jól az eredeti idősorra.

A köbös kivételével nagyon lassú monoton növekedő függvény a lineáris és a négyzetes, de igazán egyik illesztés sem használható.

# 7.8 SZEZONÁLIS DEKOMPOZÍCIÓ, MINT DETERMINISZTIKUS MATEMATIKAI MODELL

Az eddigi eredmények alapján (az eredeti időfüggvény görbéjét, vagy az autokorrelációs függvény értékeit nézve) nem várunk szezonalitást az alap idősorban, de azért elvégezhetünk egy vizsgálatot, hogy van-e mégis valamiféle periodicitás időben a jelben.

Erre dekompozíciós (más néven determinisztikus) modellt alkalmazunk, melyekkel általánosan három determinisztikus összetevő: a trendfüggvény, a szezonális ingadozásokat leíró függvény és a ciklusfüggvény adott súlyú összegét vagy szorzatát képezhetjük. Ha összegezzük ezeket a determinisztikus komponenseket, akkor additív modellről beszélünk, ha szorozzuk őket, akkor multiplikatívnak nevezzük a modellt. A három komponens mellett még számolnunk kell az úgynevezett hibataggal, mely eloszlását vizsgálnunk kell a modellillesztést követően. A hibatag egymással korrelálatlan, azonos eloszlású, nulla várható értékű, kis szórású véletlen változók sorozata, mely, ha normális eloszlást követ, akkor *fehérzaj*nak hívjuk. Amennyiben a maradéksor fehérzaj, az adott modell jól illeszkedik az idősorunkra, minden lényeges információt már megmagyaráz (13).

A dekompozíciós modellek elég egyszerű modellek. Bár nincs előfeltételük (mint az ARIMA modelleknek a stacionaritás), mégis a trendfüggvény képzésekor a mozgóátlagolás miatt adatvesztéssel jár. Mindenképpen azt feltételezi, hogy az adataink meghatározott komponensekből állnak, és a véletlen szerepét figyelmen kívül hagyja.

Az előzőek alapján a szezonalitást keressük az idősorban szezonális dekompozíciós modell futtatásával.

Ehhez először kell egy dátum változót képezni, ami jelen esetben a másodperceket fogja tartalmazni. Az SPSS-ben ez megtehető menüsorból is a  $Data \rightarrow Define \ Dates \dots$  menüponttal.

Amennyiben szintaxissal futtatjuk a következő kódrészlet első sorát írjuk be a *Syntax* ablakba (a többit az SPSS generálja a futtatással).

DATE MI 1 S 1 60. The following new variables are being created: Name Label MINUTE\_ MINUTE, not periodic SECOND\_ SECOND, period 60 DATE\_ Date. Format: "MMMM:SS"

Ennek eredményeképp létrejön három új változó az adatállományban: MINUTE\_, SECOND\_ és DATE\_, mely szerint a percen belül már a másodpercek periodicitással jelennek meg, így képezve a DATE változót.

Az adatállományban az alábbi változók jöttek létre.

📬 *m7	🚔 *m767_AH2_origin_time_sec.sav [DataSet1] - IBM SPSS Statistics Data Editor																		
<u>File E</u> d	lit ⊻i	ew	⊡ata	Transfo	rm <u>A</u>	nalyze	Direc	t <u>M</u> arket	ting G	raphs	<u>U</u> tilities Add	- <u>o</u> ns	<u>W</u> indow	Help					
		Ē	)	Ļ	<b>S</b>	1					<b>H</b>	*			\$ <u>}</u>	A		ABC	
1 : MINUT	E_		1																
		3 h14	h15	h16	h17	h18	h19	h20	h21	h22	Peer_alapId	alap_	fit_linear	fit_quad_	fit_cubic	MINUTE_	SECOND_	DATE_	var
												С	_alap	alap	_alap				1.51
1		3 17	44	16	6	8	5	4	31	14	2	6,00	7,77649	7,74965	6,59831	1	1	1:01	
2		97	2	24	31	10	21	9	43	7	1	9,00	7,77922	7,75266	6,62438	1	2	1:02	
3	3         3         27         9         7         11         18         51         17         3         35         1         7,00         7,75566         6,65025         1         3         1:03           4         2         12         27         2         4         10         12         2         400         7,78196         7,78566         6,65025         1         3         1:03																		
4	4 3 12 37 3 4 10 18 8 22 10 2 4,00 7,78469 7,75866 6,67593 1 4 1:04																		
5	;	7 4	13	42	5	10	14	16	7	25	1	8,00	7,78742	7,76166	6,70142	1	5	1:05	
6		5 11	40	5	5	17	22	16	11	6	2	4,00	7,79015	7,76465	6,72672	1	6	1:06	
7	b       p       11       4u       5       5       17       22       16       11       6       2       4,0u       7,79015       7,76465       6,72672       1       6       1:06         7       5       7       12       3       26       12       46       9       2       13       2       3,00       7,79289       7,76765       6,75182       1       7       1:07																		
8	7       7       12       3       26       12       46       9       2       13       2       3,00       7,79289       7,7656       6,75182       1       7       1:07         8       5       12       16       12       14       12       10       11       14       12       2       4,00       7,79562       7,77065       6,77673       1       8       1:08																		
9		5 17	18	15	21	8	26	16	3	19	1	7,00	7,79835	7,77364	6,80146	1	9	1:09	
10		3 31	14	7	12	13	6	16	13	7	1	10,00	7,80109	7,77664	6,82599	1	10	1:10	
11	;	7 11	44	18	4	6	32	10	16	40	2	6,00	7,80382	7,77963	6,85034	1	11	1:11	
12		7 12	14	17	10	7	4	10	26	6	2	4,00	7,80655	7,78262	6,87450	1	12	1:12	
13	;	2 9	3	10	36	7	2	12	13	17	2	6,00	7,80928	7,78562	6,89847	1	13	1:13	
14		3 7	6	5	8	10	2	13	15	12	1	12,00	7,81202	7,78861	6,92225	1	14	1:14	
15		4 5	4	6	37	22	20	35	37	19	1	7,00	7,81475	7,79160	6,94585	1	15	1:15	
r																			
Data Vi	ew	/ariabl	e View											***					

A következőkben nézzük meg a dekompozíciós modellillesztéseket.

### 7.8.1 Multiplikatív modell

A multiplikatív dekompozíciót, ha menüből kívánjuk elérni a szoftverben, akkor az *Analyze*  $\rightarrow$  *Forecasting*  $\rightarrow$  *Seasonal Decomposition*... menüpontot válasszuk. Egyébként lefuttathatjuk a következő szintaxist is.



Ahogy a szürke kommentekből is látszik, ezzel négy új változó jön létre az adatállományban, melyet a következő képen láthatunk is a *Data View* ablakban.

m767	7_AH2_	_origin_ti	ime_sec.s	av [DataSet	1] - IBM SPS	S Statisti	cs Data Editor								
<u>F</u> ile <u>E</u> dit	⊻iew	<u>D</u> ata	Transform	<u>A</u> nalyze [	Direct <u>M</u> arketing	<u>G</u> raphs	Utilities Add-ons V	<u>V</u> indow <u>H</u> elp							
			<b>,</b> k		<b>i</b>		r 🗛 🧩			-					
1 : ERR_1		,5	837308435	5495											
	ear p	fit_quad_ alap	fit_cubic _alap	MINUTE_	SECOND_	DATE_	ERR_1	SAS_1	SAF_1	STC_1	var				
1	:49	7,74965	6,59831	1	1	1:01	,58373	5,41601	1,10783	9,27827					
2	122	7,75266	6,62438	1	2	1:02	1,23809	11,03010	,81595	8,90896					
3	3         96         7,7556         6,65025         1         3         1:03         1,25830         10,28077         ,68088         8,17034           4         69         7,75566         6,67593         1         4         1:04         59973         3,55999         1,13420         6,99935														
4	-69	7,75866	6,67593	1	4	1:04	,50973	3,55809	1,12420	6,98035					
5	42	7,76166	6,70142	1	5	1:05	1,38010	8,09836	,98785	5,86796					
6	115	7,76465	6,72672	1	6	1:06	,94450	4,36046	,91733	4,61668					
7	7         89         7,76765         6,75182         1         7         1:07         53517         2,39865         1,25070         4,48203														
8	8         62         7,77065         6,77673         1         8         1:08         70871         3,91658         1,02130         5,52634														
9	35	7,77364	6,80146	1	9	1:09	1,21538	8,48982	,82452	6,98535					
10	09	7,77664	6,82599	1	10	1:10	1,57306	11,84991	,84389	7,53305					
11	82	7,77963	6,85034	1	11	1:11	,53078	3,46704	1,73058	6,53192					
12	:55	7,78262	6,87450	1	12	1:12	,64694	4,41740	,90551	6,82811					
13	128	7,78562	6,89847	1	13	1:13	,95376	7,36174	,81502	7,71865					
14	:02	7,78861	6,92225	1	14	1:14	1,64487	14,69326	,81670	8,93278					
15	<b>15</b> 75 7,79160 6,94585 1 15 1:15 <b>,77214 5,69428 1,22930 7,37468</b>														
1	4														
Data Viev	v Vari	able View													

Ezeket érdemes átnevezni az adott modell szerint, hogy további futtatások eredményei később is beazonosíthatóak legyenek. Így például adhatjuk nekik a következő változóneveket a *Variable View* ablakban.

<b>ta *</b> m767_	• m767_AH2_origin_time_sec.sav [DataSet1] - IBM SPSS Statistics Data Editor														
<u>File</u> <u>E</u> dit	<u>∨</u> iew <u>D</u> ata <u>T</u> rans	form <u>A</u> nal	yze Dire	ect <u>M</u> arketing	Graphs Utilities Add-ons Window Help										
🔁 🗄	🖨 🛄	<b>F</b>		<b>F</b>	= 📭 🌆 📰 🔛 📥 📲 🕢 🥊		86								
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role				
25	h22	Numeric	11	0		None	None	3	🔳 Right	🛷 Scale	ゝ Input	1			
26	Peer_alapId	Numeric	8	0	Peer Group ID	None	None	8	🔳 Right	\delta Nominal	ゝ Input				
27	alap_c	Numeric	8	2		None	None	3	📰 Right	🔗 Scale	🦒 Input				
28	fit_linear_alap	Numeric	11	5	Fit for alap_c from CURVEFIT, MOD_8 LINEAR	None	None	5	🔳 Right	🔗 Scale	ゝ Input				
29	29 fit_quad_alap Numeric 11 5 Fit for alap_c from CURVEFIT, MOD_9 QUADRATIC None None 5														
30	30 ft_cubic_alap Numeric 11 5 Fit for alap_c from CURVEFIT, MOD_9 CUBIC None None 5 3 Right Scale Input														
31	31 MINUTE_ Numeric 8 0 MINUTE, not periodic None 7 ≣ Right al Ordinal Nuput														
32	32 SECOND_ Numeric 2 0 SECOND, period 60 None 7 3 Right 1 Ordinal Input														
33	33 DATE_ String 7 0 Date. Format: "MMMM:SS" None None 5 ELeft & Nominal > Input														
34	34 ERR_1_mul Numeric 11 5 Error for alap_c from SEASON, MOD_1, MUL EQU 60 None 13 3 Right 🖉 Scale 🕥 Input														
35	SAS_1_mul	Numeric	11	5	Seasonal adjusted series for alap_c from SEASON, MOD_1, MUL EQU 60	None	None	13	🗮 Right	🛷 Scale	ゝ Input				
36	SAF_1_mul	Numeric	11	5	Seasonal factors for alap_c from SEASON, MOD_1, MUL EQU 60	None	None	13	🗮 Right	🔗 Scale	🦒 Input				
37	STC_1_mul	Numeric	11	5	Trend-cycle for alap_c from SEASON, MOD_1, MUL EQU 60	None	None	13	🔳 Right	🔗 Scale	ゝ Input				
38															
39															
	۲ ۲ ۲														
Data View	Data View Variable View														
							IE	BM SPSS Sta	tistics Proces	sor is ready					

Ezen változók jelentése a következő:

ERR\_1\_mul: ez a hibatag / maradéktag (a generált név az error szóra utal)

**SAS\_1\_mul**: ebben a változóban tárolódik a szezonális hatástól megfosztott idősor. Ez a korrigált alap jel és a szezonális tag hányadosa: alap\_c / SAF\_1\_mul (a "seasonally adjusted" kifejezésből ered az elnevezés)

SAF\_1\_mul: ez a szezonális tag, ami a "seasonal factor" elnevezésből eredeztethető

**STC\_1\_mul**: ezen változó tartalmazza a trend hatást. Neve a "trend-cycle" kifejezésből ered. A szezonális hatástól megfosztott idősor simított változata, vagyis a SAS\_1\_mul változóból simítással készül.

Ha ezeket az eredeti (korrigált alap) idősorral együtt megjelenítjük akár menüsorból (*Analyze*  $\rightarrow$  *Forecasting*  $\rightarrow$  *Sequence Charts...*), akár a következő kódrészlet lefuttatásával, akkor láthatjuk a különálló hatásokat.

\*Ezeket jelenitsuk meg egyutt az eredeti idosorral. **TSPLOT VARIABLES=alap\_c err\_1\_mul sas\_1\_mul saf\_1\_mul stc\_1\_mul** /ID=DATE\_ /NOLOG /FORMAT NOFILL NOREFERENCE.

A kapott eredmény a következő.



Bár eléggé átfedik egymást a görbék, de annyi látható, hogy a hibatag és a szezonális komponens együtt mozog, valamint külön, de hasonló nagyságrendben az eredeti alap\_c függvény, a trendfüggvény és a szezonális komponenstől megfosztott idősor. Ezt így is várjuk.

Külön-külön megvizsgálva az egyes komponenseket, hogy jobban lássuk és értsük a folyamatot, nézzük meg először csak a szezonális összetevőt.



A fenti szintaxis futtatásának eredménye a következő diagram.



Az ábráról leolvasható, hogy minden perc 10-edik másodperce környékén van egy csúcs, ami eltér az 1-től 1,75 körüli értékkel. Ez egy **periodicitást** mutat. Látható továbbá, hogy a szezonális komponens a minimum értékeket az adott perc 30-adik másodperce környékén veszi föl, míg a maximumot az előbb említett 10 sec. körül.

A multiplikatív modellben az 1 faktor felel meg a szezonalitás hiányának. A szezonális komponens minimum értéke megközelítőleg 0,62, míg maximuma 1,75 körül van.

Amennyiben csak a trendkomponenst nézzük, az alábbi kódsort lefuttathatjuk.

\*csak a trendkomponens az eredeti korrigalt alappal egyutt.
TSPLOT VARIABLES=alap\_c stc\_1\_mul
/ID=DATE\_
/NOLOG
/FORMAT NOFILL NOREFERENCE.



A kapott eredmény diagram jól mutatja a trendfüggvény korrigált alap jelre való illeszkedését.

A trendfüggvényben nem igazan látszódik hosszú távú ciklus, hullámzás, leginkább random váltakozik a görbe.

A maradéksort is nézzük meg külön. Itt ellenőriznünk kell a normalitást is. Ehhez először a következő programkóddal szemléltessük a hibatagot.



A futtatás eredménye az alábbi.



Láthatóan az 1 körül oszcillál a maradéksor, és semmilyen felismerhető mintázatot nem mutat, stacionáriusnak tűnik.

A normális eloszlás ellenőrzésére rajzoljunk ki egy P-P diagramot, majd hisztogramot és végül egy Kolmogorov-Smirnov próbát is számoljunk az alábbi szintaxis szerint.



Az eredmények a következők.





A P-P diagramon az átlótól egy szakaszon eltérnek a pontok, egy másfajta (mint lineáris) szabályszerűséget mutatnak. A hisztogram enyhén jobbra ferde eloszlást mutat két csúccsal. A Kolmogorov-Smirnov próba 0,966 átlagú, 0,43 szórású eloszlást mutat, mely 5%-os szinten szignifikánsak eltér a normális eloszlástól (p=0,011). Itt is megjegyezném, hogy független adatok esetén értelmezett, ugyanakkor gazdasági elemzésekben és az SPSS könyvben is ezt alkalmazzák, ezért mi is lefuttatjuk, de fenntartásokkal kezeljük.

Ezek alapján tehát az 1 várható értékű normális eloszlástól szignifikánsan eltér a **hibatag**, azaz, mint idősor **fehérzajnak nem tekinthető**. Eszerint **az idősor multiplikatív dekompozíciója nem mondható sikeresnek**.

Ellenőrzésként hozzuk létre a multiplikatív modell változóját az

képlettel a modellünkből visszaszámolva.

A következő kódsorral létrehozzuk előbb ezt a változót (COMPUTE), majd együtt az eredeti alap\_c idősorral, majd külön-külön is kirajzoljuk a TSPLOT utasításokkal.

COMPUTE seasonal\_decomp\_multipl\_model=SAS\_1\_mul\*SAF\_1\_mul\*ERR\_1\_mul. EXECUTE. TSPLOT VARIABLES=alap\_c seasonal\_decomp\_multipl\_model /NOLOG /FORMAT NOFILL NOREFERENCE. \*kulo-kulon. \* Sequence Charts. TSPLOT VARIABLES=alap\_c /ID=time /NOLOG /FORMAT NOFILL NOREFERENCE. TSPLOT VARIABLES=seasonal\_decomp\_multipl\_model /ID=time /NOLOG /FORMAT NOFILL NOREFERENCE.

A COMPUTE hatására az adatállományban létrejött a *seasonal\_decomp\_multipl\_model* elnevezésű változó, mely a modell illesztett értékeit tartalmazza.

Azon az ábrán, ahol az *alap\_c* és a *seasonal\_decomp\_multipl\_model* idősorok is szerepelnek, ez utóbbi meglehetősen elfedi az alapot, ezért néztük meg külön-külön is.



Külön a korrigált alap idősor.



Majd a szezonális dekompozíciós modellillesztés idősora.



A két ábrát összehasonlítva, értékeiket figyelembe véve elmondható, hogy a két görbe gyakorlati szempontból nem egyezik meg. A szezonális dekompozíciós görbe nagyobb értékeket mutat.

### 7.8.2 Additív modell

Az additív szezonális dekompozíciós modellt az előző mintájára végezhetjük el. Ugyancsak az *Analyze*  $\rightarrow$  *Forecasting*  $\rightarrow$  *Seasonal Decomposition*... menüből érhető el, vagy a következő parancssorral.



A kódban is látszik, hogy a MODEL paraméterben kell most az ADDITIVE kulcsszót megadni a korábbi MULTIPLICATIVE helyett. Ezzel is négy új modellfüggvény változó jött létre az adatállományban, melyeket érdemes az előző mintájára elnevezni (most az additív modell miatt \_*add* véggel):

<b>ħ *</b> m767_J	*m767_AH2_origin_time_sec.sav [DataSet1] - IBM SPSS Statistics Data Editor													
<u>File Edit y</u>	/iew <u>D</u> ata <u>T</u> ransform <u>A</u> nalyze Direct	Marketing	<u>G</u> raphs <u>l</u>	<u>J</u> tilities Ad	d-ons Window Help									
) 😂 🔓	🖨 🛄 🗠 🛥 📱	1		<u> </u>	📓 📓 📥 📥 📲 🎝 🌑 🤲									
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measur				
34	ERR_1_mul	Numeric	11	5	Error for alap_c from SEASON, MOD_1, MUL EQU 60	None	None	13	🚎 Right	🛷 Scale 🖆				
35	SAS_1_mul	Numeric	11	5	Seasonal adjusted series for alap_c from SEASON, MOD_1, MUL EQU 60	None	None	13	🗏 Right	🛷 Scale				
36	SAF_1_mul	Numeric	11	5	Seasonal factors for alap_c from SEASON, MOD_1, MUL EQU 60	None	None	13	🗏 Right	🛷 Scale				
37	STC_1_mul	Numeric	11	5	Trend-cycle for alap_c from SEASON, MOD_1, MUL EQU 60	None	None	13	📰 Right	🛷 Scale				
38	seasonal_decomp_multipl_model	Numeric	8	2		None	None	31	🗏 Right	🔗 Scale				
39	ERR_1_add	Numeric	11	5	Error for alap_c from SEASON, MOD_10, ADD EQU 60	None	None	13	🗃 Right	🛷 Scale				
40	SAS_1_add	Numeric	11	5	Seasonal adjusted series for alap_c from SEASON, MOD_10, ADD EQU 60	None	None	13	🗮 Right	🛷 Scale				
41	SAF_1_add	Numeric	11	5	Seasonal factors for alap_c from SEASON, MOD_10, ADD EQU 60	None	None	13	🗮 Right	🛷 Scale 🚽				
42	STC_1_add	Numeric	11	5	Trend-cycle for alap_c from SEASON, MOD_10, ADD EQU 60	None	None	13	🗮 Right	🛷 Scale				
43														
44														
45														
46														
47														
48														
49														
50														
51	51													
	4									4				
Data View	Variable View													
					IBA	4 SPSS Stat	istics Proces	sor is ready						

ERR\_1\_add: hibatag/maradéktag

SAS\_1\_add: a szezonális hatástól megfosztott idősor (alap\_c - SAF\_1\_add)

SAF\_1\_add: szezonális tag

STC\_1\_add: trendfüggvény

Az eredeti korrigált alap idősorral együtt megjelenítve az új változókat a következő szintaxis lefuttatásával kaphatjuk meg a grafikont.



Itt is kirajzolódik, amit a multiplikatív modell esetén láthattunk, hogy a hibatag és a szezonális komponens kb. egy skálán mozog, míg az alap\_c idősorral együtt megy a szezonális komponenstől megfosztott összetevő idősora, és az ebből simítással nyert trendfüggvény.

Ha külön-külön megnézzük, előbb a szezonális komponenst, ismét TSPLOT utasításokat kell felparamétereznünk.





Látható, hogy a szezonális komponens a minimum értékeket az adott perc 30-adik másodperce környékén veszi föl (min. ~ -3), míg a maximumot 10 sec. körül (max. ~ 6), ugyanúgy, mint a multiplikatív modellben, csak itt a 0-tól való eltérést nézzük, és más a skála.

Az additív modellben a 0 faktor felel meg a szezonalitás hiányának. Eszerint itt is megfigyelhető egyfajta periodicitás.

Ha külön csak a trendkomponenst nézzük a korrigált alapra vetítve, láthatjuk az illeszkedést.

\*csak a trendkomponens az eredeti korrigalt alappal egyutt.
TSPLOT VARIABLES=alap\_c stc\_1\_add
/ID=DATE\_
/NOLOG
/FORMAT NOFILL NOREFERENCE.

A fenti kódsor lefuttatásával kapjuk a következő ábrát.



A trendfüggvényben nem igazán látszódik hosszú távú ciklus, hullámzás, vagy egyenletesen növekedő tendencia (talán egy nagyon enyhe emelkedés csak), leginkább random váltakozik a görbe.

Ellenőrizzük, hogy a maradéksor fehérzaj folyamat-e. Ehhez előbb jelenítsük meg (TSPLOT), majd nézzük meg az előzőek mintájára a P-P diagramot (PPLOT) és a Kolmogorov-Smirnov próba (NPTESTS) eredményét. Ezekhez a következő szintaxisokat futtassuk le az SPSS-ben.



A nemparaméteres próbához a normális eloszlástól való eltérés vizsgálatához pedig:

```
*Nonparametric Tests: One Sample.

NPTESTS

/ONESAMPLE TEST (ERR_1_add) KOLMOGOROV_SMIRNOV(NORMAL=SAMPLE)

/MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE

/CRITERIA ALPHA=0.05 CILEVEL=95.
```

A kapott eredmények a következők lettek.





Normal P-P Plot of Error for alap\_c from SEASON, MOD\_10, ADD EQU 60

Már az idősor diagramon is látható, hogy az adatok random szóródnak a nulla körül.

A P-P diagramon a pontok kb. követik az átlót, de egy kis eltérő szabályosság is megfigyelhető (egy nagyon enyhe hullám - más függvényszerű kapcsolat, nem random szóródás).

A Kolmogorov-Smirnov próba 0,112-es p-értéket adott, miszerint a maradéktag nem különbözik szignifikánsan (5%-os szinten) a fehérzajtól (véletlenszerű változás a *sequence plot* alapján, normális eloszlás a PPLOT es NPTESTS alapján). Ezért az idősor additív dekompozíciója sikeresnek tekinthető, bar sokat nem árul el. (Azért itt is figyeljünk arra, hogy a normalitás teszt független adatokra működik leginkább.)

Ellenőrizhetjük most is az additív modellünket. Ehhez képezzük a következő rövid szintaxis lefuttatásával a *seasonal\_decomp\_add\_model* változót.

COMPUTE seasonal\_decomp\_add\_model=SAS\_1\_add+SAF\_1\_add+ERR\_1\_add. EXECUTE.

Megjelenítve az illesztett modellt és az eredeti korrigált alap idősort az alábbi kódot futtathatjuk, amivel egy diagramon, de külön-külön is kirajzoljuk a grafikonokat.

```
    TSPLOT VARIABLES=alap_c seasonal_decomp_add_model
/NOLOG
    /FORMAT NOFILL NOREFERENCE.
    *kulo-kulon.
    * Sequence Charts.
    TSPLOT VARIABLES=alap_c
/ID=time
/NOLOG
    /FORMAT NOFILL NOREFERENCE.
    TSPLOT VARIABLES=seasonal_decomp_add_model
/ID=time
/NOLOG
    /FORMAT NOFILL NOREFERENCE.
```

Az eredményt a következő három ábra mutatja, melyeken jól látható, hogy azért az additív modell is jócskán elfedi a korrigált alap idősor értékeit, vagyis az illesztés nem használható igazán, hiába találtuk fehérzaj folyamatnak a maradéktagot. Eszerint ennél többet viszont ilyen idősor elemzéssel nehéz kinyerni az adatokból.



A korrigált alap idősor görbéje:



Az illesztett additív szezonális dekompozíciós modell.



A kapott grafikonokon jól látható, hogy a két görbe gyakorlati szempontból nem egyezik meg. A szezonális dekompozíciós görbe nagyobb értékeket mutat a multiplikatívhoz hasonlóan, mint a korrigált alap idősor.

### 7.9 EXPONENCIÁLIS SIMÍTÁS

A simító eljárások a determinisztikus modelleknél jobban figyelembe veszik az idősor véletlen jellegét, belső összefüggéseit. A véletlen hatást simítással (kiátlagolással) próbálják csökkenteni. Az exponenciális szűrés egyike ezen simító eljárásoknak, mely során determinisztikus modellből kiindulva lépésenként előrejelzünk (csak egy lépéssel becsülünk előre), majd a keletkező hibataggal negatív visszacsatolással korrigálunk. A módszer neve onnan ered, hogy az idősor t-edik időpillanatban vett elemét az őt megelőző (múltbeli) elemek exponenciálisan csökkenő súlyokkal vett lineáris kombinációjával becsüljük (13).

Az exponenciális simítás módszerét először 1956-ban alkalmazta Robert Goodell Brown (14).

Az IBM SPSS szoftverben természetesen az előzőek mintájára ezt az elemzést is el lehet érni menüből. Ehhez az *Analyze*  $\rightarrow$  *Forecasting*  $\rightarrow$  *Create Models*... menüpontot válasszuk, ahol *Method*-nak állítsuk be az *Exponential Smoothing* opciót. Ha parancssorból szeretnénk futtatni, akkor a következő szintaxist használhatjuk föl.



Az utolsó sorban az EXSMOOTH opcióban láthatjuk, hogy a Simple típusú modellt futtattuk, mely esetben semmilyen trendet, vagy szezonális hatást nem tételezünk fel az idősorban.

Az eredményből jól látható, hogy például a modell magyarázó erejét leíró **R-négyzet** érték **0-hoz közeli** (lásd *Model Fit* vagy *Model Statistics* táblákban az *R-squared* értékét), ami azt mutatja, hogy az **illesztés korrelálatlan az idősortól**. Ez tehát megint csak egy rosszul illeszkedő modell az eredeti (korrigált) idősorra. (Ugyanakkor itt is figyelnünk kell arra, hogy az R-négyzet független adatok esetén használható igazán a modell magyarázó erejének leírására, itt viszont idősorról beszélünk, mely összetartozó adatokat tartalmaz!)

A *Residual ACF* és *Residual PACF* táblákból valamint a generált diagramokból is látható, hogy a maradéktag (hibatag, reziduál értékek idősora) autokorrelálatlan idősor, egészen a 24-edik tagig visszamenve (LAG értéke 24 az AUXILIARY utasítás MAXACFLAGS opciójában), hiszen alacsonyak az autokorrelációs és parciális autokorrelációs értékek (a korrelogramokon a

konfidencia sávot nem is lépik túl). Ezt alátámasztja a Model Statistics táblában található Ljung-Box tesztstatisztika p-értéke (p=0,818), ami ugyancsak jelzi, hogy a maradéktag fehérzajnak minősíthető. Ettől függetlenül maga a függvény nem modellezi az idősorunkat.

A Model Statistics tábla eredményét, majd a hibatag autokorrelációs és parciáis autokorrelációs ábráit láthatjuk a következő képeken, végül magát az illesztett exponenciális szűrési modell eredményét idősor grafikonon ábrázolva az eredeti korrigált alap (alap\_c) idősorral együtt szemléltetve.



Az exponenciálisan szűrt idősort láthatjuk (kék) a következő ábrán a megfigyelt értékekre (piros) vetítve 95%-os konfidencia intervallummal együtt.



Ha megnézzük a Simple módszer helyett a Holt-féle lineáris trend illesztéssel kapott exponenciális simító eljárást, akkor ahhoz a következő szintaxist kell lefuttatnunk.

\*Exponential smoothing - Holt's linear trenddel.
 \*TSMODEL
 /MODELSUMMARY PRINT=[MODELFIT]
 /MODELSTATISTICS DISPLAY=YES MODELFIT=[ SRSQUARE RSQUARE]
 /MODELDETAILS PRINT=[ PARAMETERS RESIDACF RESIDPACF FORECASTS] PLOT=[ RESIDACF RESIDPACF]
 /SERIESPLOT OBSERVED FORECAST FIT FORECASTCI FITCI
 /OUTPUTFILTER DISPLAY=ALLMODELS
 /SAVE PREDICTED(Predicted) NRESIDUAL(NResidual)
 /AUXILIARY CILEVEL=95 MAXACFLAGS=24
 /MISSING USERMISSING=EXCLUDE
 /MODEL DEPENDENT=alap\_c
 PREFIX='Model'
 /EXSMOOTH TYPE=HOLT TRANSFORM=NONE.

Itt az utolsó sorban az EXSMOOTH utasítás típusaként a HOLT kulcsszót adhatjuk meg. (A Holt és Simple módszerek hasonló eredményt adnak, a Holt módszer több helyen használt.)

A Holt típusú exponenciális simító modell R-négyzet értéke is igen alacsony (-0,043), ami hasonlóan az előző Simple modellhez azt jelzi, hogy az **illesztés** itt is **korrelálatlan az idősortól**. A Ljung-Box próba p-értéke itt már 0,097, ami bár 5 %-os szinten nem szignifikáns, de 10%-os  $\alpha$  mellett már igen, vagyis 10%-os szinten már szignifikánsan eltér a maradéktag a fehérzajtól. Így megállapíthatjuk, hogy ez sem jó modell az adataink leírására, bár nem is nagyon vártuk az eddigi eredmények alapján (előző Simple modell, vagy a sima lineáris trend illesztés).

A kapott eredmények most az alábbiak.



Date

Az adatállományban lementésre kerültek a SAVE utasítás hatására a megfelelő illesztett függvények és a második esetben, a Holt-féle lineáris trendnél a reziduálok is. Ezeknek a változó neveit és címkéit érdemes beszédesebben elnevezni mondjuk az alábbi kép szerint a *Variable View* ablakban.

幅 *m767_A	m767_AH2_origin_time_sec.sav (DataSet1) - IBM SPSS Statistics Data Editor													
<u>File E</u> dit <u>V</u>	jew <u>D</u> ata <u>T</u> ransform <u>A</u> nalyze Direct <u>M</u> a	rketing <u>G</u> raphs	<u>U</u> tilities A	dd- <u>o</u> ns <u>Wi</u> r	ndow <u>H</u> elp									
) 😂 🔓	🖨 🗳 🖕 🛥 📓		н н		🔟 🧰 🛃 📲 📲 🖉 🍋									
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns						
31	MINUTE_	Numeric	8	0	MINUTE, not periodic	None	None	7	<b>1</b>					
32	SECOND_	Numeric	2	0	SECOND, period 60	None	None	7	畫					
33	DATE_	String	7	0	Date. Format: "MMMM:SS"	None	None	5	E.					
34	ERR_1_mul	Numeric	11	5	Error for alap_c from SEASON, MOD_1, MUL EQU 60	None	None	8	遭					
35	SAS_1_mul	Numeric	11	5	Seasonal adjusted series for alap_c from SEASON, MOD_1, MUL EQU	None	None	8	1					
36	SAF_1_mul	Numeric	11	5	Seasonal factors for alap_c from SEASON, MOD_1, MUL EQU 60	None	None	8	遭					
37	STC_1_mul	Numeric	11	5	Trend-cycle for alap_c from SEASON, MOD_1, MUL EQU 60	None	None	8	<b>=</b>					
38	seasonal_decomp_multipl_model	Numeric	8	2		None	None	22	-					
39	ERR_1_add	Numeric	11	5	Error for alap_c from SEASON, MOD_10, ADD EQU 60	None	None	8	Ξ.					
40	SAS_1_add	Numeric	11	5	Seasonal adjusted series for alap_c from SEASON, MOD_10, ADD EQ	None	None	8	-					
41	SAF_1_add	Numeric	11	5	Seasonal factors for alap_c from SEASON, MOD_10, ADD EQU 60	None	None	8	=					
42	STC_1_add	Numeric	11	5	Trend-cycle for alap_c from SEASON, MOD_10, ADD EQU 60	None	None	8	=					
43	seasonal_decomp_add_model	Numeric	8	2		None	None	22	-					
44	Predicted_alap_c_Simple_ExpSmooth	Numeric	8	2	Predicted value from alap_c - Exponential Smoothing - Simple method	None	None	8	=					
45	Predicted_alap_c_Holt_ExpSmooth	Numeric	8	2	Predicted value from alap_c - Exponential Smoothing - Holt's linear trend	None	None	8	-					
46	NResidual_alap_c_Holt_ExpSmooth	Numeric	8	2	Noise residual from alap_c - Exponential Smoothing - Holt's linear trend	None	None	8	=					
47														
48									-					
	1					-								
Data View	/ariable View													

### 7.10 ARIMA MODELLILLESZTÉS

A legárnyaltabb, legösszetettebb idősoros elemzés a **Box** és **Jenkins** által kidolgozott ún. ARIMA modellekkel lehetséges. A szóösszetétel az AR, I és MA részekre bontható, melyek az autoregresszív, integrált és mozgóátlag folyamatok elnevezéseire utalnak. Ezek a modellek az idősor adatai között meglévő belső sztochasztikus koherenciát feltételeznek, melyek tartósan jelen vannak az idősorban, ami által pontos előrejelzések várhatók a modellekben.

Az **autoregresszív folyamatok**ban egy adott t időpontbeli érték az őt megelőző (múltbeli) értékek lineáris kombinációjaként (súlyozott összegeként) és egy korrelálatlan hibatag összegeként áll elő.

A **mozgóátlagolás** során egy végtelen hosszú fehérzaj folyamat egyfajta mozgóátlaga és egy korrelálatlan hibatag összegeként modellezzük az idősort.

Beszélhetünk ún. **ARMA** modellekről is, melyekben az idősor egy autoregresszív, egy mozgóátlag és egy korrelálatlan hibatag összegeként definiálható. Az integrált ARIMA modellek esetén az idősor *deriváltsora* (szomszédos elemek különbsége) lesz ARMA típusú (*13*).

Az IBM SPSS-ben ugyancsak az *Analyze*  $\rightarrow$  *Forecasting*  $\rightarrow$  *Create Models*... menüpontban érhető el az ARIMA modellillesztés is. Itt választhatjuk az Expert Modeler opciót is, amivel a 19-es verziótól kezdve már lehetőségünk van automatikus modellillesztésre, vagyis a szoftver megkeresi, hogy melyik illeszkedik a legjobban. Ehhez az EXPERTMODELER utasítást használhatjuk, ahogy az alábbi szintaxisban is olvasható.



Az eredmény, vagyis a legjobban illeszkedő ARIMA vagy exponenciális simító modell egy kb. konstans modell lett (8,53-as értékkel mindenhol, egy kivétellel, ami 4:02-kor 48), semmilyen transzformációt nem végzett. Eszerint nem nagyon lehet mit kihozni így ebből az idősorból. Az *R-squared* érték is 0,084, ami azt jelenti, hogy mindössze 8,4%-ban magyarázza ez a konstans modell az idősor teljes varianciáját (független adatokat feltételezve!).

Az illesztett függvény változója létrejött a fenti szintaxis lefuttatásával az adatállományban, amit érdemes a korábban látottak mintájára átnevezni a megfelelő modellre utalva.

🍓 m767_	AH2_origin_	time_sec.sav	[DataSet1] - IBM SPSS Statistics D	ata Editor										
<u>F</u> ile <u>E</u> dit	<u>∨</u> iew <u>D</u> ata	<u>T</u> ransform <u>A</u>	Analyze Direct <u>M</u> arketing <u>G</u> raphs U	tilities Add-ons	: <u>Wi</u> ndow <u>H</u>	lelp								
		<b>II.</b> 🗠	Image: A state of the state	•	5			<b>ABG</b>						
1 : Predicted	_alap_c_Con	8,529215378707	766					Visibl	e: 47 of 47 Vari	iables				
	∿F_1_add	STC_1_add	seasonal_decomp_add_model	Predicted_ala p_c_Simple_ ExpSmooth	Predicted_ala p_c_Holt_Ex pSmooth	NResidual_al ap_c_Holt_E xpSmooth	Predicted_alap_c_Constant	var	var					
1	1,39689	7,12317	3,48	7,90	6,39	- ,39	8,53							
2	2,28023	6,94508	8,77	7,88	6,36	2,64	8,53							
3	-2,51237	6,58891	9,92	7,89	6,61	,39	8,53							
4	1,27652	5,97204	,75	7,88	6,65	-2,65	8,53							
5	,72652	5,33089	9,94	7,85	6,40	1,60	8,53							
6	-1,28644	4,24735	5,04	7,85	6,56	-2,56	8,53							
7	2,37467	4,02903	- ,40	7,81	6,32	-3,32	8,53							
8	,15430	5,29035	2,56	7,77	6,00	-2,00	8,53							
9	9         -1,84755         6,52965         9,32         7,73         5,81         1,19         8,53													
10	-1,84385	6,92101	14,92	7,73	5,93	4,07	8,53							
11	5,78023	5,66566	,55	7,75	6,32	- ,32	8,53							
12	- ,77718	6,24920	2,53	7,73	6,29	-2,29	8,53							
13	-2,24200	43809, 7	6,80	7,70	6,08	- ,08	8,53							
14	-1,14385	8,61093	16,53	7,68	6,07	5,93	8,53							
15	,84504	6,97924	6,18	7,72	6,64	,36	8,53			Ţ				
·	1													
Data View	Variable Viev	v												
							IBM SPSS Statistics Processo	r is ready						

Címkének adhatjuk a Variable View nézetben például a "Predicted value from alap\_c - Expert Modeler best fit: Constant model" leírást.

Az Output ablakban generált eredményekből a következők jól összegzik a lényeget.

		Moc	lel Statistics	6			
Model	Number of	Model Fit s	tatistics	Ljung	-Box Q(1	8)	Number of
	Predictors	Stationary R-	R-squared	Statistics	DF	Sig.	Outliers
		squared					
alap_c-Model_1	0	,084	,084	13,505	18	<mark>,761</mark>	1

A Ljung-Box próba alapján a maradéksor fehérzajnak tekinthető, vagyis a modellünk már minden lényeges információt megmagyaráz.

### ARIMA Model Parameters

				Estimate	SE	t	Sig.
alap_c-Model_1	alap_c	- No Transformation	Constant	8,529	,217	39,254	,000

Az *ARIMA Model Parameters* táblában látható, hogy a transzformáció nélküli modell lett szignifikánsan illeszkedő az alap\_c idősorunkra. Ez tulajdonképpen az ARIMA(0,0,0) modell.

	Outliers													
			Estimate	SE	t	Sig.								
alap_c-Model_1	4:02	Additive	39,471	5,322	7,416	,000								

Az *Outliers* táblázat mutatja, hogy az idősorban egy elem van (4:02-kor), ami kiugró, vagyis jelentősen eltér ebben az időpontban a modellből számított érték az eredeti mért értéktől. Ez az, amit korábban már tapasztaltunk a generált változó vizsgálata során az adatállományban.

A Residual ACF és PACF grafikonok alátámasztják a Ljung-Box próba eredményét.



Az illesztett modell és az eredeti korrigált alap idősor grafikonját láthatjuk végül.



Date

### 7.11 SPEKTRÁL ELEMZÉS

A sztochasztikus idősor elemzésnek két ágát különíthetjük el: Az egyik az eddigiekben is taglalt időtartományon alapuló vizsgálat, amely az idősor dinamikájára, az egymást követő megfigyelésekre épít sztochasztikus idősori modellek segítségével (ARIMA modellek). A másik terület a frekvenciatartományban történő elemzés, amely a ciklikus komponensek súlyának összességét vizsgálja a spektrálanalízis módszerével.

A ciklikus folyamatokban a periódus hossza legtöbbször ismert és a szezonális komponensben figyelembe van véve. Előfordulhatnak azonban a természeti folyamatokban is rejtett ciklusok az idősorban, melyek kimutatására alkalmas egyik statisztikai módszer a **periodogram**.

A periodogram matematikai oldalról közelítve a gyors **Fourier transzformáció**n alapszik (FFT: Fast Fourier Transform), mellyel az időtartományról áttérünk frekvencia tartományra szinusz és koszinusz függvények kombinációjaként.

A periodogram megadható periódus vagy frekvencia függvényében is (ezek egymás reciprokai: f=1/T). Amikor periódus szerint adjuk meg, értéke annál a periódusnál nagy, amely szerint az idősorban a ciklikusság tapasztalható. Frekvencia alapon történő ábrázolás esetén, ha valamely frekvenciánál kiugró értéket észlelünk, létezik a folyamatnak spektrális összetevője. Akkor nevezzük inkább periodogramnak, ha periódus függvényében ábrázoljuk. A frekvenciák függvényében történő ábrázolást inkább **spektrogram**nak hívjuk. A **periodogram simításával kapjuk a spektrogramot**: súlyozással spektrális ablakot (ablakfüggvényt) használnak. Különféle ablakfüggvények léteznek, mint a Hamming-Tukey, a Parzen vagy például a Bartlett módszer. Ez utóbbi módszer gyakran alkalmazott a variancia probléma csökkentésére (adott frekvencián a variancia emelkedő esetszám ellenére sem csökken). A módszer lényege abban rejlik, hogy az eredetileg n elemű idősort részekre osztjuk, melyekben külön-külön diszkrét Fourier transzformációt végzünk, majd a négyzetüket átlagoljuk (*13, 15*).

Ha lenne a szezonális dekompozíció eredményeként kapott trendkomponensben valamiféle hullámzás, hosszú távú ciklus, érdemes lenne periodicitás elemzést végezni az eredeti korrigált adatsoron.

Bár most ilyet nem tapasztaltunk, nézzük meg azért a spektrum elemzés egyik módszerének gyakorlása végett.

IBM SPSS-ben az Analyze  $\rightarrow$  Forecasting  $\rightarrow$  Spectral Analysis... menüben érhető el a spektrum elemzés.

Ha szintaxisból szeretnénk futtatni, az alábbi kódsort használhatjuk.



A Bartlett-ablak került felhasználásra, mint az látható a WINDOW utasításban.

Az eredmény periodogram a következő.



Periodogram of alap\_c by Period

A kapott periodogramon látható, hogy sehol nem vesz fel kimagasló maximum értéket, vagyis nem tapasztalható ismétlődés ilyen másodperces adatokat figyelembe véve. Random szóródnak az értékek.

Frekvencia alapon vizsgálódva az alábbiképpen írhatjuk át a szintaxisunk (csak a PLOT utasítás paraméterei változnak).

```
frekvencia alapon nezzuk, es a spectral density plot-ot is bekattintjuk.
 Spectral Analysis.
TSET PRINT=DEFAULT.
    CTRA
S
  VARIABLES=alap_c
    NDOW=BARTLETT(5)
   ENTER
 /PLOT=P S BY FREQUENCY.
```

A spectral density (spektrális sűrűségfüggvény vagy más néven spektrum) függvény a periodogram simított változata. Ez már mentes az esetleges ingadozásoktól.

A periodogram itt a következőképpen néz ki már a frekvenciatartományon vizsgálódva.



Periodogram of alap\_c by Frequency

Az eredmények között lekértük a Spectral Density diagramot, azaz a spektrumot is.



Window: Bartlett (5)

Itt is látható, hogy random szóródnak az értékek (random váltják egymást a csúcsok és völgyek), nem érzékelhető periodicitás ebből.

Ha a parciális korrelogram alapján már igazoltunk volna valamilyen periodicitást, akkor azt ellenőrizhetnénk itt is a megfelelő Fourier-frekvencia kiszámításával. Ha pl. 20 másodpercenként lenne valami a jelben, akkor az 1/20 értéknél kellene kb. lennie az első nagy csúcsnak a spektrum függvényben. Ezt követően szabályosan ismétlődhetne (2/20, 3/20, stb. értékeknél újabb csúcsokkal). Ha így több csúcs lenne, akkor azt mutatná, hogy a jel nem tisztán szinuszoid idősor (mert a tisztán szinuszoid idősorok spektrumában van egyetlen csúcs csupán).

De nincs ilyen csúcsunk sehol, vagyis újabb bizonyíték, hogy nincs szezonalitás az alap korrigált adatsorunkban.

# 8 Köszönetnyilvánítás

Nagyon köszönöm Prof. Bari Ferencnek a konstruktív javaslatokat és hasznos megbeszéléseket a problémák megértése érdekében. Hálával tartozok Dr. Domoki Ferencnek és Tóth-Szűki Valériának a kísérlet, a mintavétel és a szakmai célok érthető összefoglalásáért, egyeztetésekért és korrekciókért, valamint Oláh Orsolyával együtt a kísérlet pontos lebonyolításáért. Prof. Hantos Zoltánnak köszönöm szépen a szakmai támogatást és útbaigazítást.

Végtelen hálás vagyok Lang Zsoltnak, akihez szakmai kérdésekkel tudtam fordulni az adatelemzés és kapott eredmények értelmezése kapcsán, melyekhez nagyon hasznos tanácsokat adott.

# 9 Irodalomjegyzék

1. B. Hajtman, *Bevezetés a biostatisztikába: nem csak orvosoknak* (Edge 2000 Kiadó, Budapest, 2012), pp. 242. p.: ill.; 21 cm.

2. S. A. Book, L. K. Bustad, The fetal and neonatal pig in biomedical research. J. Anim. Sci. 38, 997-1002 (1974).

3. P. A. Flecknell, R. Wootton, M. John, Total body glucose metabolism in the conscious, unrestrained piglet and its relation to body- and organ weight. *Br. J. Nutr.* **44**, 193-203 (1980).

4. O. Olah, V. Toth-Szuki, P. Temesvari, F. Bari, F. Domoki, Delayed Neurovascular Dysfunction Is Alleviated by Hydrogen in Asphyxiated Newborn Pigs. *Neonatology*. **104**, 79-86 (2013).

5. F. Bari, T. M. Louis, W. Meng, D. W. Busija, Global ischemia impairs ATP-sensitive K+ channel function in cerebral arterioles in piglets. *Stroke*. **27**, 1874-80; discussion 1880-1 (1996).

6. D. W. Busija *et al.*, Effects of ischemia on cerebrovascular responses to N-methyl-D-aspartate in piglets. *Am. J. Physiol.* **270**, H1225-30 (1996).

7. F. Domoki *et al.*, Hydrogen is neuroprotective and preserves cerebrovascular reactivity in asphyxiated newborn pigs. *Pediatr. Res.* **68**, 387-392 (2010).

8. S. Kéri, B. Gulyás, Elektrofiziológiai módszerek a kognitív idegtudományokban. **5. fejezet**, 81-96.

9. L. Détári, Biológiai jelek számítógépes elemzése. (1983).

10. L. Schnell, G. Blahó, in (Tankvk.; Műegyetemi K, Budapest, 2001), pp. 345 p. ;.

11. Statistical Analysis System Institute, *SAS certification prep guide: base programming for SAS 9* (SAS Inst, Cary, NC, ed. 3, 2011).

12. Keh-Dong Shiang, The SAS® Calculations of Areas Under the Curve (AUC) for Multiple Metabolic Readings. , 1-14 .

13. L. Ketskeméty, L. Izsó, T. Könyves Előd, Bevezetés az IBM SPSS Statistics programrendszerbe: módszertani útmutató és feladatgyűjtemény statisztikai elemzésekhez (Artéria Stúdió, Budapest, ed. 3. jav., átd. kiad., 2011).

14. R. G. Brown, Exponential smoothing for predicting demand. 15 Pages.(1956).

15. J. Pintér, A spektrálanalízisről. Statisztikai Szemle. 85. évfolyam, 130-156 (2007).

# 10 Támogatás

A kutatás a TÁMOP 4.2.4.A/2-11-1-2012-0001 azonosító számú Nemzeti Kiválóság Program – Hazai hallgatói, illetve kutatói személyi támogatást biztosító rendszer kidolgozása és működtetése országos program című kiemelt projekt keretében zajlott. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.