

# Bevezetés a SAS Enterprise Miner 5.3 használatába

Készítette: Soltész Gábor

[solteszgabee@gmail.com](mailto:solteszgabee@gmail.com)

2010.



| Új Calculus Bt.

Cím: 1132 Budapest, Visegrádi u. 30., Levelezési cím: 1397 Budapest 62. Pf. 550.

E-mail: [office@calculus.hu](mailto:office@calculus.hu), Web: [www.calculus.hu](http://www.calculus.hu), Tel./fax: +36/1/463-2274, Mobil: +36/20/935-0645, 405-6684

## Tartalom

<b>1. SAS Enterprise Miner 4.x projekt átalakítása 5.3-as változatba .....</b>	<b>4</b>
Példa: .....	5
<b>2. Kezelőfelület.....</b>	<b>6</b>
<b>3. Új projekt létrehozása.....</b>	<b>7</b>
<b>4. Fájlok importálása .....</b>	<b>7</b>
<b>5. Diagram létrehozása .....</b>	<b>11</b>
5.1. SAS kód alkalmazása.....	13
5.2. Leíró statisztika készítése .....	15
5.3. Hiányzó adatok pótlása .....	17
5.3.1. Csere csomópont használata (Replacement) .....	18
5.4. Új szabály létrehozása .....	19
5.5. Új változó létrehozása .....	22
5.6. Szűrő (Filter) alkalmazása .....	25
5.7. Multiplot használata.....	26
5.8. Elrendezés megváltoztatása .....	26

## Ábrajegyzék

1. ábra Kezelőfelület	6
2. ábra Importálási beállítások	7
3. ábra Adatforrás meghatározása	9
4. ábra Explorer ablak	9
5. ábra Importált SAS adattábla	10
6. ábra Diagram létrehozása	11
7. ábra Adatforrás elhelyezése a munkaterületen	11
8. ábra Adatforrás szerkesztése	12
9. ábra Információ az adatforrásról	12
10. ábra Explore funkció	12
11. ábra Adott csoport elemeinek kiemelése	13
12. ábra SAS csomópont hozzáadása	13
13. ábra SAS kód csomópont beállítások	14
14. ábra Csomópont frissítés	14
15. ábra Módosítás utáni hisztogram	15
16. ábra StatExplore beállítások	15
17. ábra Statisztikai eredmények	16
18. ábra Eredmények táblázatos formában	16
19. ábra Hiányzó adatok pótlása	17
20. ábra Hiányzó adatok pótlásának beállításai	17
21. ábra Csere csomópont alkalmazása	18
22. ábra A pótlás és a csere két változata	18
23. ábra Folyamatábra (szabályalkotóval)	19
24. ábra Szabály megadása	19
25. ábra Szabályalkotó oldal	20
26. ábra Szabályfa alkotás	20
27. ábra További bontás	20
28. ábra A szabály eredménye	21
29. ábra Folyamatábra	22
30. ábra Beállítási lehetőségek	22
31. ábra Változó létrehozása	23
32. ábra Kapott eredmény	23
33. ábra A generált változó eloszlása	24
34. ábra Szűrő csomópont elhelyezése	25
35. ábra Szűrési kritérium	25
36. ábra A teljes folyamatábra	26
37. ábra Multiplot beállítási lehetőségei	26

## 1. SAS Enterprise Miner 4.x projekt átalakítása 5.3-as változatba

Az 5.3-as változat lehetőséget biztosít a 4.x változattal készített adatprojektek megfelelő környezetbe történő áthelyezésére. Ezt az átalakítást elvégezve lehetőség van a korábbi folyamatábrák, eredmények, kimenetek megtartására.

Az átalakítást a %EM\_CONVERT makró végzi.

**%EM\_CONVERT**(

**EM4PROJECT**=*SAS-Enterprise-Miner-4.x-project-path*,

**PROJECTPATH**=*SAS-Enterprise-Miner-5.3-project-path*,

**PROJECTNAME**=*SAS-Enterprise-Miner-5.3-project-name*,

**CREATE**=Y|N (*generate-Enterprise-Miner-5.3-project-folders-with-results*),

**FILEREF**=*fileref-for-Enterprise-Miner-5.3-project-creation-batch-code*);

**EM4PROJECT** a SAS Enterprise Miner 4.x projekt elérési útja

**PROJECTPATH** az 5.3-as változat projekt elérési útja

**PROJECTNAME** a projekt neve

**CREATE** (opcionális) alapértelmezés szerint Yes. Arra utal, hogy szeretnénk e projekt könyvtárstruktúrárt létrehozni, ami tartalmazza a folyamatábrákat, meta adatokat és az eredményeket. Ahhoz, hogy 4.x-ből 5.3-as változatba konvertáljuk szükséges a könyvtárstruktúra létrehozása. A könyvtárstruktúra létrehozása helyett lehetőség van egy generáló batch kód létrehozására, ehhez a CREATE=N feltétel szükséges, továbbá a következő parancsot is ki kell tölteni.

**FILEREF (opcionális)** a generált batch kód tárolására alkalmas könyvtár elérési útja és a keletkezett fájl neve. Ha a könyvtárstruktúrárt és a generáló kódot is létre szeretném hozni, akkor az előző pontban a CREATE=Y utasítás mellett ezt a pontot is ki kell tölteni.

Konvertálás lépései:

1. El kell dönteni, hogy a könyvtárstruktúrárt és a generáló kódot is létre akarjuk e hozni.
2. A kívánt 4.x-es projektről a következő információk szükségesek:
  - A konvertálni kívánt projekt elérési útja
  - Az elérési utat, ahova az új projektet menteni akarom, vagy a generáló kódot létre akarom hozni.
  - A konvertálni kívánt projekt neve
3. Indítsuk el a SAS 9.1.3 (vagy későbbi) szoftvert (SAS Enterprise Miner 4.3 nem futthat) és futtassuk a konvertáló makrót.
4. A makró futása után indítsunk el a SAS Enterprise Miner 5.3-t. Majd hozzunk létre egy új projektet.

- Host: azt a SAS Enterprise Miner 5.3 szerveret kell kiválasztani, ahol a konvertált projekt található.
  - Path: A szerver elérési útja.
  - Name: A névnek meg kell egyeznie a konvertált projekt könyvtár nevével.
5. Ha a konvertálás sikeres volt, akkor a Project Navigatorban a Diagram könyvtárat lenyitva meg kell, hogy jelenjenek a konvertált folyamatábrák.

### Példa:

\$user: a felhasználó/fiók neved

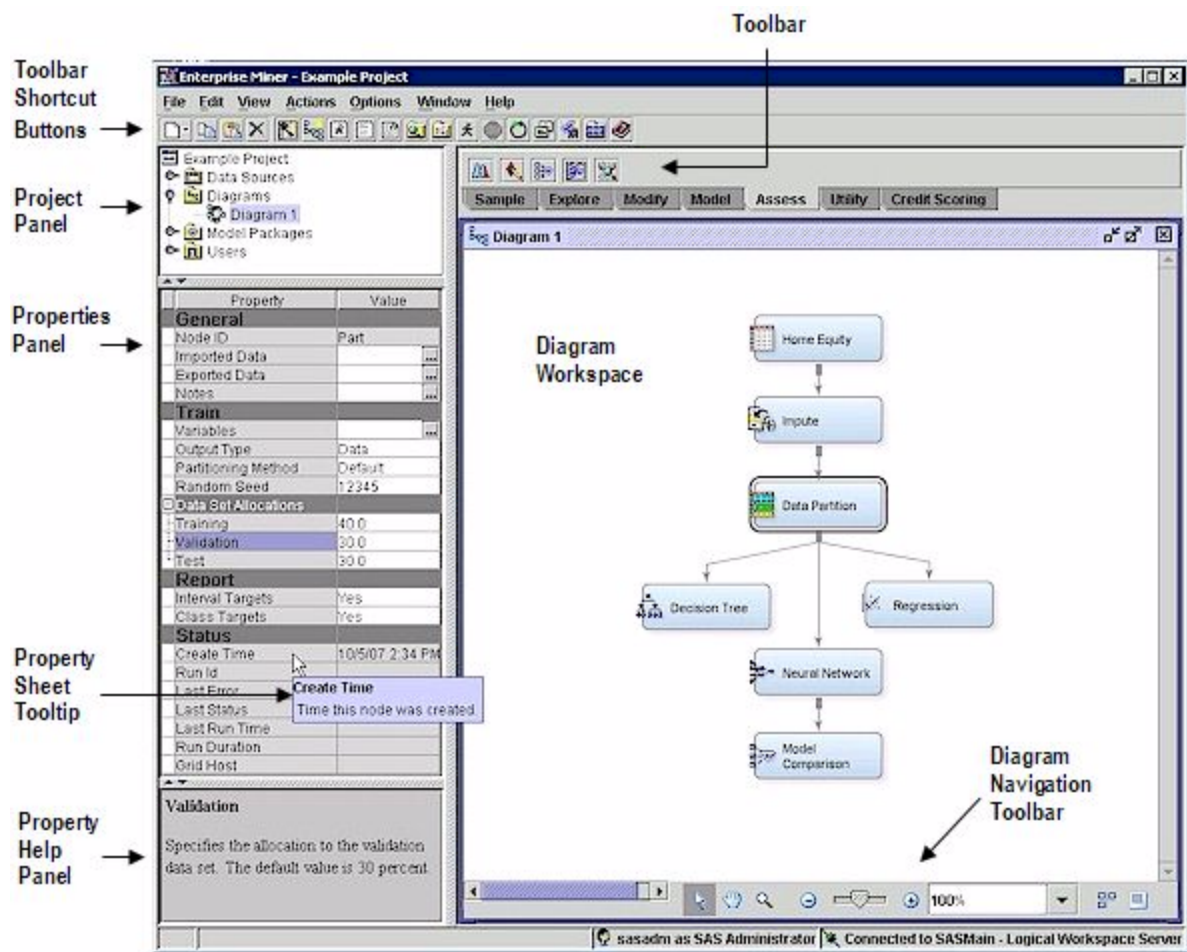
\$projekt: a konvertálandó projekt neve

```
%let em4project=C:\Documents and Settings\$user\Dokumentumok\My SAS
Files\9.1\EM Projects\$projekt;
%let em5path=C:\Documents and Settings\$user\Dokumentumok\My SAS
Files\9.1\EM_Projects\;
%let em5name=$projekt;

%em_convert(em4project=&em4project, projectPath=&em5path,
            projectName=&em5name);

%macro em_convert(em4project=, projectPath=, projectName=,fileref=, create=Y);
proc display c=sashelp.emconvrt.convert.scl;
Futtatás;
quit;
%mend em_convert;
```

## 2. Kezelőfelület



1. ábra Kezelőfelület

**Eszköztár parancsikoni gombok:** Eszközsor parancsikon / alapvető funkciók (létrehozás, mentés, futtatás...)

**Projekt panel:** Ha több folyamatábrát és forrásállományt akarunk használni, akkor itt könnyen rendszerezhetjük, és áttekinthetjük projektünk tartalmát

**Tulajdonságok panel:** Minden csomóponthoz tartozik egy ilyen táblázat, aminek a segítségével a csomópont teszte szabható.

**Tulajdonság lap:** Általános információt tartalmaz minden csomópontról.

**Tulajdonság leíró panel:** Az egyes funkciókról tartalmaz rövid leírást.

**Diagram munkaterület:** Munkaterület, a folyamatábra, folyamatábrák ezen helyezkednek el.

**Diagram navigátor eszköztár:** A megjelenítést segítő eszközök.

### 3. Új projekt létrehozása

A SAS Enterprise Miner 5.3 elindítása után.

#### **File / New / Project**

**Name:** a létrehozandó projekt neve

**Host:** a kívánt szerver amin létrehozni és futtatni kívánjuk a projektet.

**Path:** a projekt elérési útja.

**Start-Up Code / Exit code:** az itt megadott SAS kód a projekt futtatásakor / bezárásakor végrehajtódik.

### 4. Fájlok importálása

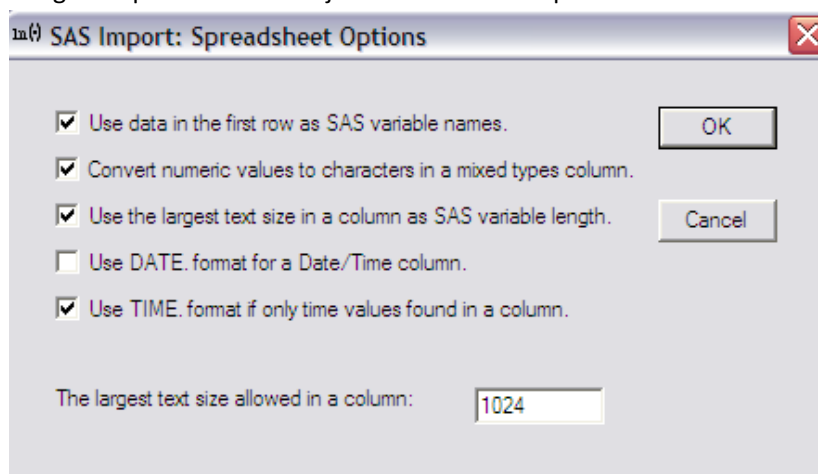
Az importálandó fájl: *kekszalag2008\_erre\_jav.xls*

Mivel ez egy excel fájl, így az Enterprise Miner nem képes importálni a fájlt. Így az importálás előtt mindenképpen szükséges SAS adattáblát generálni az excel fájlból.

- Első lépésként tekintsük meg az excel fájlt. Jól látható, hogy a táblázat felett találhatóak bejegyzések. Ezeket feltétlen törölni kell, mert a SAS az importáláskor csak a legfelső sort képes megkülönböztetni, hiszen adott esetben ez tartalmazza az oszlopok neveit. Töröljük a bejegyzéseket és mentjük el a fájlt.
- Indítsuk el a SAS Base-t.
- File / Import Data...

Itt válasszuk ki a **Microsoft Excel 97, 2000 or 2002 Workbook** típust.

Keressük meg az importálni kívánt fájlt és állítsuk be az opciókat az alábbiak szerint.



2. ábra Importálási beállítások

- Adjuk meg a generálandó fájl könyvtárát és nevét.

A következő lépésben adatmanipulációt hajtunk végre a létrehozott fájlon.

A következő SAS kód azért szükséges, hogy eltávolítsuk a felesleges oszlopokat, mivel az excel táblázatban bizonyos sorok formátummal szerepeltek, így a SAS úgy importálta a fájlt, hogy az üres oszlopokat is hozzávette. Így szükséges leválogatni es beszédesebb változónevet adni.

```
libname em 'c:\em';

data em.keksz_EM(keep=F1 F2 F3 hajonev ys kormányos Befutasi_ido Futott_ido
Korrigalt_ido);
  set work.em;
  rename haj_n_v=hajonev
         korm_nyos = kormányos
         Befut_si_id_____pp_ss_ = befutasi_ido
         Futott_id____sec___ = futott_ido
         Korrig_lt_id____sec___ = korrigalt_ido;
Run;
```

/\*Az alábbi kód azért szükséges, hogy a befutási időt átkonvertáljuk numerikus típusúvá\*/

```
data em.keksz_em;
  set em.keksz_em;
  hr=scan(befutasi_ido,1, '.');
  min=scan(befutasi_ido,2, '.');
  sec=scan(befutasi_ido,3, '.');
  befutasi_ido_sec=hms(hr,min,sec);
  drop hr min sec befutasi_ido;
  label befutasi_ido_sec='Befutási idő (sec)';
Run;
```

Ezek után létrehoztunk egy sas „EM” nevű könyvtárat, mely a c:\em könyvtárra mutat és ebben található meg a *keksz\_EM* adattábla. Ez már alkalmas arra, hogy importáljuk az Enterprise Miner-be.

Ezek után elindítjuk a SAS Enterprise Miner 5.3 szoftvert.

Létre kell hoznunk egy új projektet. Ezt az alábbiak szerint tehetjük meg.

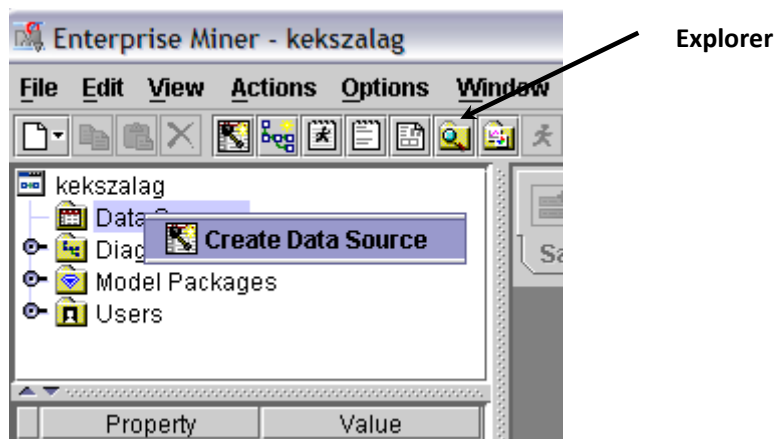
### **Fájl / New / Project**

Meg kell adnunk a projekt nevét és a SAS szervert, amin dolgozni akaFuttatásk.

Esetemben a projekt neve **kekszalag** a szerver pedig a **SASMain – Logical Workspace server**

A projekt létrehozása után az első legfontosabb dolog a forrásállomány(ok) definiálása. Ezt a project panel segítségével egyszerűen megtehetem az *Adatforrásra* (DataSource) ikonra kattintva. (3. ábra)



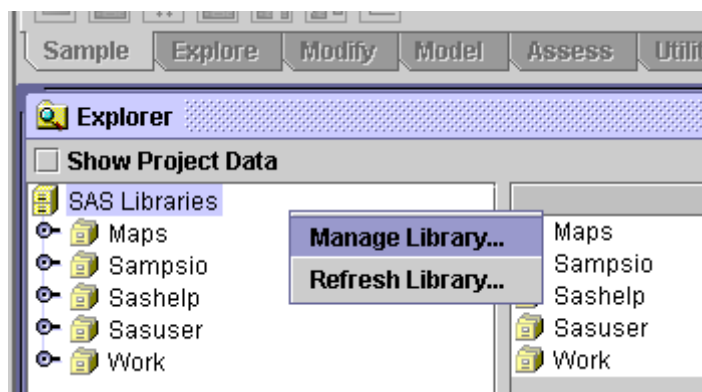


3. ábra Adatforrás meghatározása

Metaadat forrásnak válasszunk SAS táblát.

Következő lépés a kívánt SAS tábla kiválasztása. Azonban itt alapértelmezés szerint csak a programmal telepített és definiált könyvtárak találhatók, így első lépésként egy új könyvtárat kell létrehozni. Ezt a 3. ábra alapján az *Intéző (Explorer)* segítségével tehetjük meg.

Az Explorer ablakban jobb klikk majd a *Könyvtárkezelő (Manage Library...)* (4. ábra)



4. ábra Explorer ablak

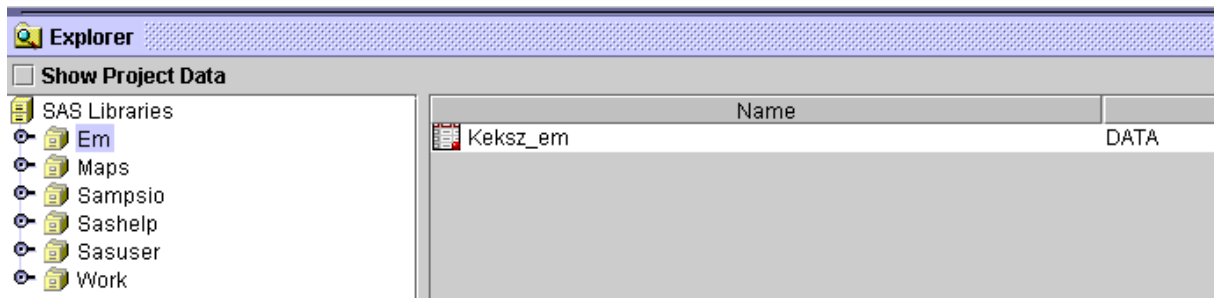
Új könyvtárat akarunk létrehozni. Ehhez definiálnunk kell a könyvtár nevét és az adott könyvtár elérési útját, esetemben ez *EM*, és az elérési út *c:\EM*.

**FONTOS**, hogy annak a könyvtárnak az elérési útját adjuk meg, amiben a korábban létrehozott SAS állomány található (*keksz\_em.sas7bdat*). Ellenkező esetben csak azokkal a SAS táblákkal fogunk tudni dolgozni, amik az adott könyvtárban találhatók.

Amint létrehoztuk a kívánt könyvtárat ellenőrizzük, hogy a kívánt névvel és tartalommal hoztuk e létre, ehhez először is frissíteni kell a könyvtárstruktúrát.

Klikkeljünk a *Könyvtárfrissítés (Refresh Library...)* parancsra. (4. ábra)

Ha mindent jól csináltunk, akkor a következő ábrát kell, hogy kapjuk. (5. ábra)



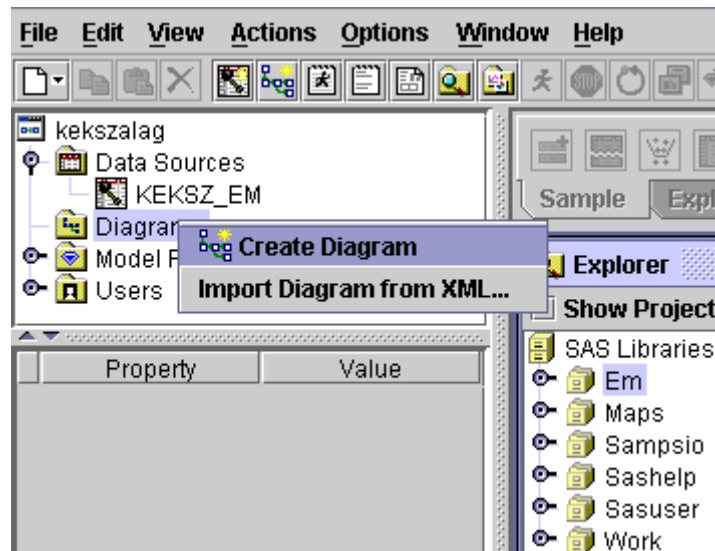
5. ábra Importált SAS adattábla

Ezek után hozzuk létre az adatforrást, amivel dolgozni szeretnénk, ehhez használjuk az *Adatforrás létrehozás (Create Data Source)* utasítást használva. (3. ábra) Forrásnak válasszuk ki az *Em* könyvtárban található *Kecszen* SAS adattáblát. Ezek után a projekt panel *Adatforrások* pontja alatt megjelenik a kiválasztott tábla.

## 5. Diagram létrehozása

Ahhoz, hogy adatbányászati folyamatokat hozzunk létre, szükséges létrehozzunk egy diagramot (Diagram munkaterület - 1. ábra), amire az egyes folyamatokat le tudjuk képezni.

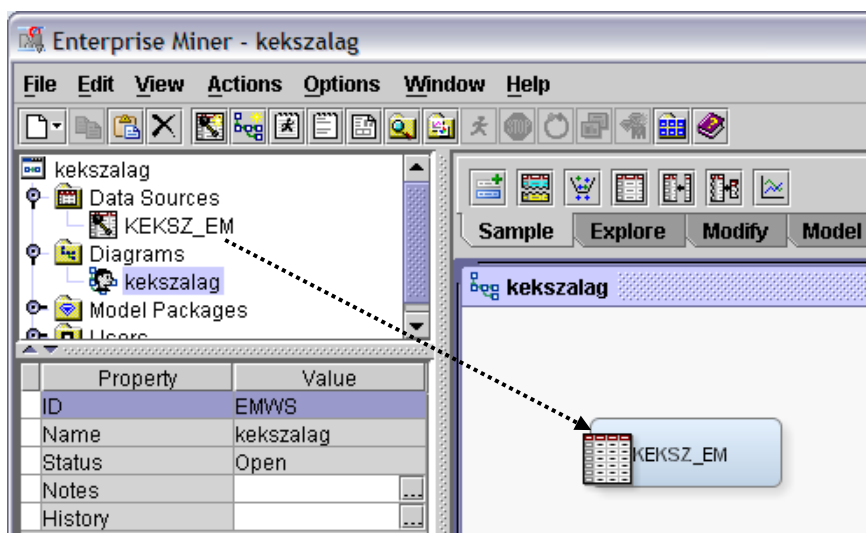
Diagramot a Project Panel segítségével tudunk létrehozni a *Diagramok* mappára klikkelve, majd a *Diagram létrehozása* parancsot választva. (6. ábra) A létrehozott diagram neve legyen *kekszalag*



6. ábra Diagram létrehozása

Ezek után létrejön a *Munkaterület* (Workspace), ahova a továbbiakban a kívánt csomópontokat el tudjuk helyezni, ezekből munkafolyamokat tudunk alakítani és végül ezeket futtatva ki tudjuk értékelni az eredményeket.

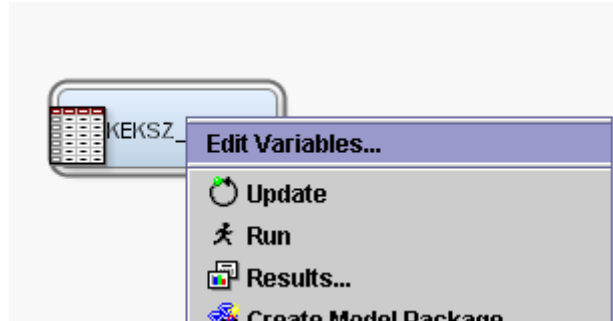
Minden folyam első lépése, amiből ki kell indulni egy SAS adattábla. Az előző pontban definiált adatforrást, ami az *Adatforrás* mappában található, át kell húzni (Drag & Drop) a munkaterületre. (7. ábra)



7. ábra Adatforrás elhelyezése a munkaterületen

Most, hogy a munkaterületen található az adatforrás a következő lépés a benne található változók megismerése. A legegyszerűbb vizsgálatot az *Adatforrás* csomópontban tehetjük.

Klikkeljünk jobb egérgombbal a munkaterületen található csomópontunkra, majd válasszuk a *Változó szerkesztése* parancsot. (8. ábra)



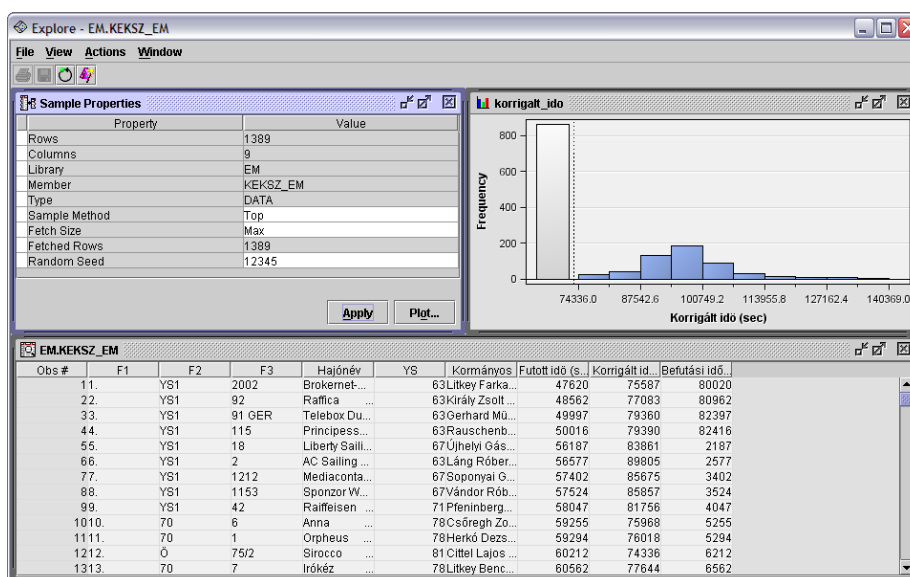
8. ábra Adatforrás szerkesztése

Ezek után megjelenik egy táblázat, mely a forrásfájlban található változókról tartalmaz információt.

Name	Role	Level	Report	Order	Drop	Lower	Upper Limit	Type	Label	Format	Informat	Length
F1	Input	Nominal	No		No			Character	F1	\$4.0	\$4.0	4.0
F2	Input	Nominal	No		No			Character	F2	\$4.0	\$4.0	4.0
F3	Input	Nominal	No		No			Character	F3	\$8.0	\$8.0	8.0
hajonev	Input	Nominal	No		No			Character	Hajónév	\$23.0	\$23.0	23.0
kormanyos	Input	Nominal	No		No			Character	Kormányos	\$27.0	\$27.0	27.0
YS	Input	Interval	No		No			Numeric	YS			8.0
befutasi_ido	Input	Interval	No		No			Numeric	Befutási idő (sec)			8.0
futott_ido	Input	Interval	No		No			Numeric	Futott idő (sec)			8.0
korrigalt_ido	Input	Interval	No		No			Numeric	Korrigált idő (sec)			8.0

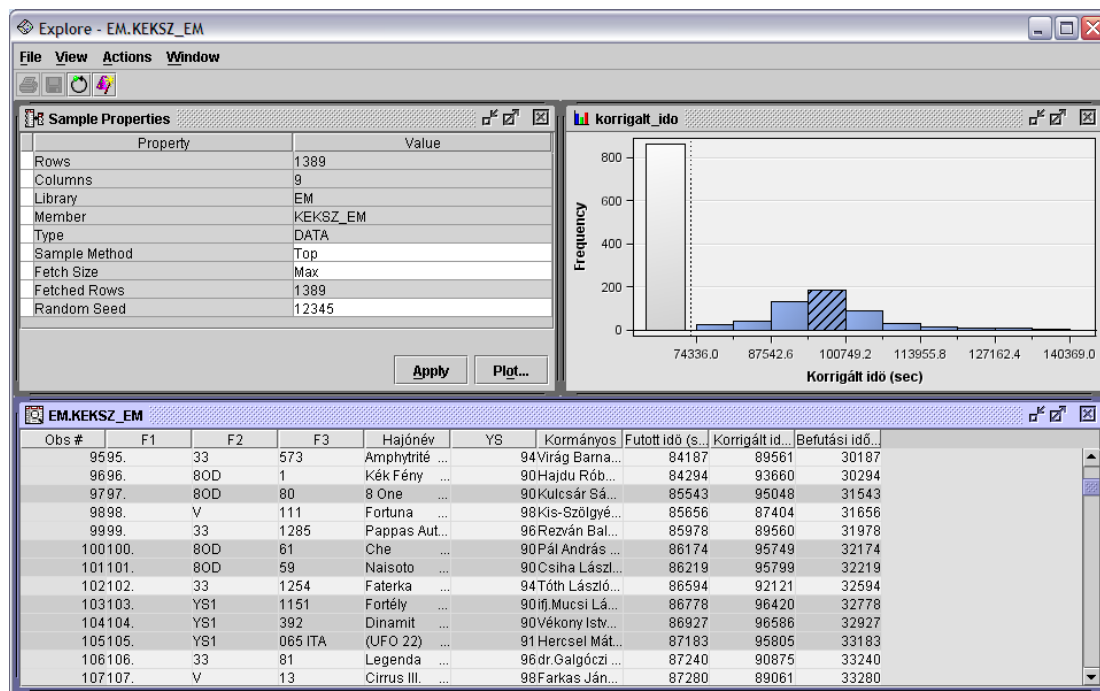
9. ábra Információ az adatforrásról

A táblázatból jól látható, hogy 5 karakter és 4 numerikus típusú változót tartalmaz az adatállomány. Válasszuk ki az utolsó (korrigalt\_ido) változót, majd nyomjuk meg a *Feltár (Explore...)* gombot.



10. ábra Explore funkció

Ekkor megjelenik a következő képernyő. (10. ábra) A „Korrigált idő” változó eloszlását figyelhetjük meg. Lényeges előrelépés az előző verzióhoz képest, hogy az 5.3-as változatban lehetőség van arra, hogy a hisztogramon bizonyos oszlopot kiválasztva, az abba tartozó elemeket kiemeljük. (11. ábra)

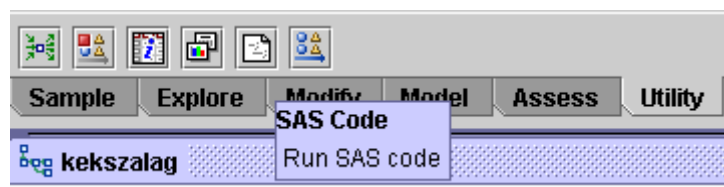


11. ábra Adott csoport elemeinek kiemelése

Az ábrából azonban észrevehetjük, hogy vannak olyan sorok, amihez nem tartozik érték. Ezt megfigyelhetjük, ha a hisztogram alatt található listát a végére görgetjük. Itt találunk üres sorokat. Ezeket feltétlenül kezelni kell, mert befolyásolják az eredményt. A legnagyobb problémát az ebbe az oszlopba tartozó sorok elemszáma jelenti, hiszen 865 sor esetében nincs információnk. Ez soknak mondható, tekintve, hogy 575 ismert sorunk van.

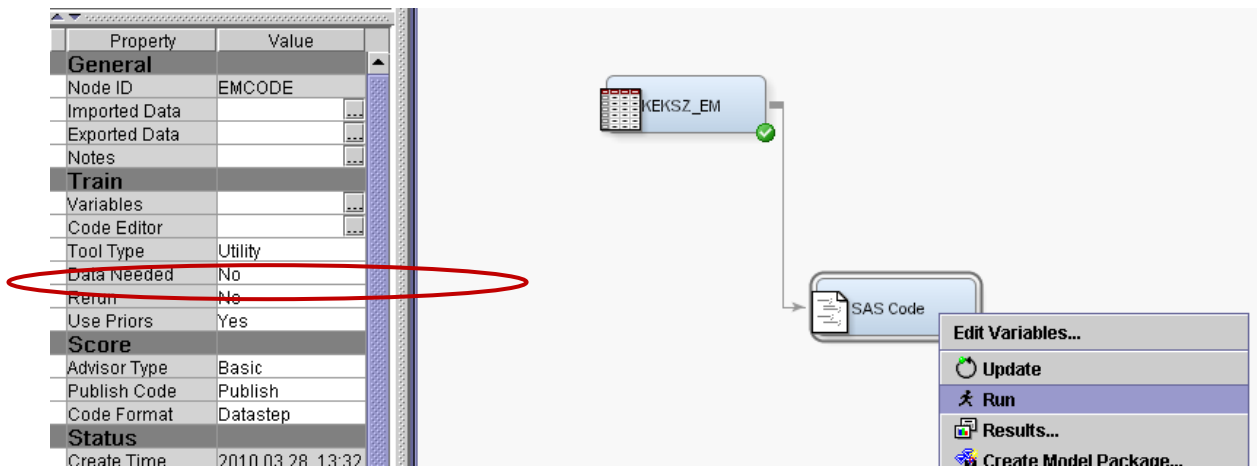
### 5.1. SAS kód alkalmazása

Ahhoz, hogy módosítsuk az adatállományunkat, SAS kódot kell futtatnunk Enterprise Miner alatt. Erre a munkaterület felett a *Segédalkalmazás (Utility)* fülön a **SAS kód** csomópontot kell a folyamunkhoz kapcsolni. (12. ábra)



12. ábra SAS csomópont hozzáadása

A következő ábra szemlélteti, hogy a csomópontot a forráshoz kapcsolva, milyen beállításokat tehetünk. Válasszuk a **Kód szerkesztés** parancsot a kívánt SAS kód beírásához. (13. ábra)



13. ábra SAS kód csomópont beállítások

A Training Code ablakba gépeljük be az alábbi kódot.

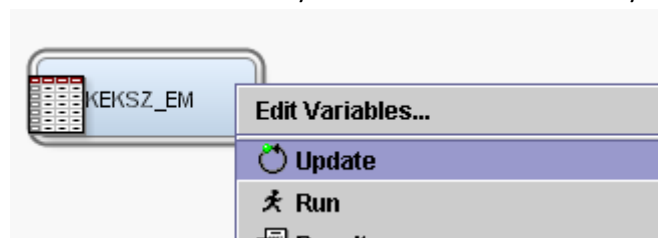
```
libname em 'c:\em';
```

```
data em.keksz_em;
  set em.keksz_em;
  where hajonev ^= ";
```

```
Run;
```

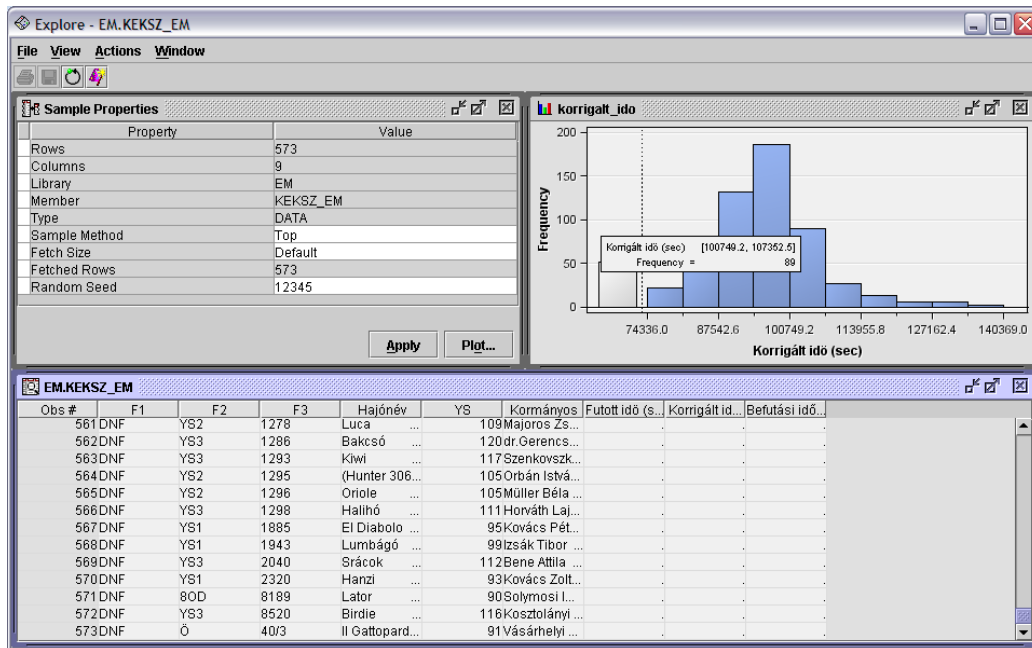
Majd a felső menüsorban lévő *Mentés* ikonra kattintva mentjük el a kódot és zárjuk be az ablakot. Ezek után a 13. ábra jobb oldalán látható *SAS kód* csomópontra klikkelve futtassuk a beírt kódot a *Futtatás* utasítás segítségével.

Következő lépésben frissíteni kell a forrásállományt. Jobb klikk a forrásállományon, majd *Frissítés*.



14. ábra Csomópont frissítés

Ha a továbbiakban a 8. ábra szerint folytatjuk a vizsgálatainkat, akkor egy letisztultabb hisztogramot kapunk, melyben a hiányzó értékek száma mindössze 51 és ezek a sorok tartalmaznak karakteres értékeket és „csak” a numerikus értékek hiányoznak. (15. ábra)



15. ábra Módosítás utáni histogram

## 5.2. Leíró statisztika készítése

A histogramon kívül további számszerűsíthető értékek is szükségesek az átfogóbb elemzésekhez, ebben nyújt támogatást számunkra a *StatFeltárás (StatExplore)* csomópont. Ez az elem a *Feltárás* fülön található. (Kezelőfelület - 1. ábra)

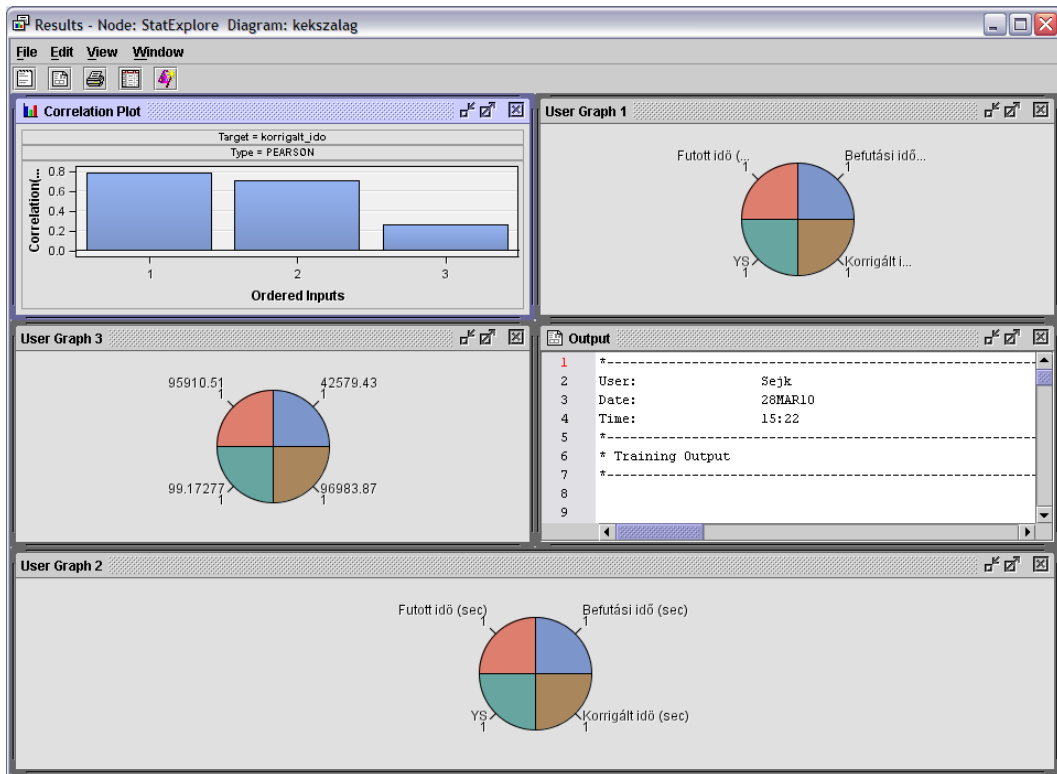
Első lépésként ki kell választani egy célváltozót, amire szeretnénk a leíró statisztikai grafikonokat elkészíteni. A célváltozót az adatforrás állományban *Változó szerkesztése (Edit Variable...)* a *Szabály* oszlopban lehet beállítani. Az adott változó esetén a szabályt *célváltozóra (Target)* állítva.

Az adatforrás után kapcsolnunk kell egy *StatFeltárás* csomópontot. Itt be kell állítani a következőket.

Property	Value
<b>General</b>	
Node ID	Stat
Imported Data	...
Exported Data	...
Notes	...
<b>Train</b>	
Variables	...
Use Segment Vari	No
Variable Selection	
Hide Rejected Vari	Yes
Number of Selecte	1000
Chi-Square Statisti	
Chi-Square	Yes
Interval Variables	Yes
Number of Bins	2
Correlation Statisti	
Correlations	Yes
Pearson Correlatic	Yes
Spearman Correla	No
<b>Status</b>	
Create Time	2010.03.28. 14:46
Run Id	9d1ea2e4-f373-42

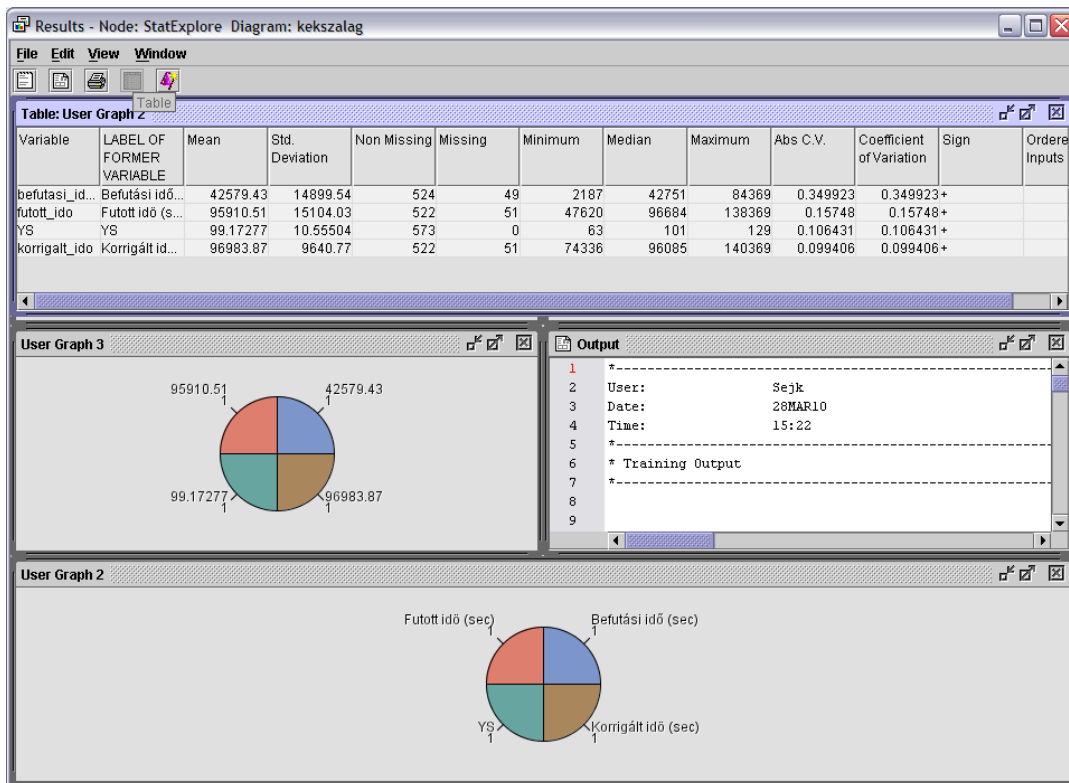
16. ábra StatExplore beállítások

A beállítások után futtassuk a csomópontot a **Futtatás** paranccsal. Az eredményeket a 17. ábra mutatja.



17. ábra Statisztikai eredmények

Számszerű eredményeket a menüsorban található *Tábla* ikon segítségével kérhetjük le. (18. ábra)

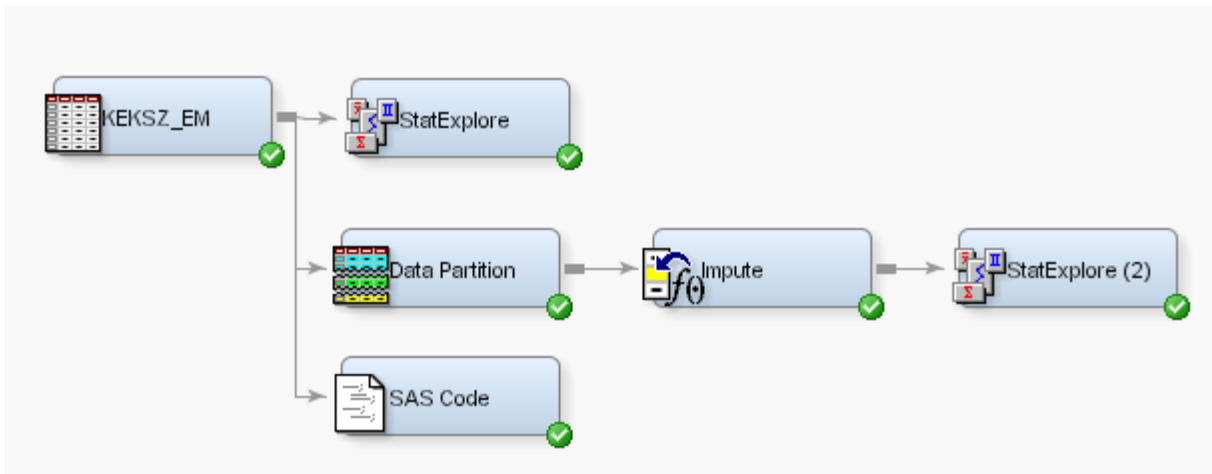


18. ábra Eredmények táblázatos formában



### 5.3. Hiányzó adatok pótlása

A hiányzó adatok pótlásához az *Pótlás (Impute)* csomópont használatos, mely a *Módosítás (Modify)* eszköztárban található. Kössük ezt a csomópontot a 19. ábra szerint.



19. ábra Hiányzó adatok pótlása

A 20. ábra alapján jól megfigyelhető, hogy az adatokat típusainak megfelelően kezeli a Miner:

- osztályozó változók (class variables)
- intervallum típusú változók (interval variables)

Ezt tovább bonthatjuk aszerint, hogy célváltozóval dolgozunk e, mert azokra különböző pótlási lehetőséget adhatunk meg.

Esetemben csak intervallum típusú változók vannak, így csak a bekarikázott rész fontos számomra.

Az intervallum típusú változókat átlag alapján pótolom, míg a célváltozót eloszlás alapján.

Property	Value
<b>General</b>	
Node ID	Impt
Imported Data	...
Exported Data	...
Notes	...
<b>Train</b>	
Variables	...
Non Missing Variables	No
Missing Cutoff	50.0
Class Variables	
Default Input Method	Default Constant Value
Default Target Method	Distribution
Normalize Values	Yes
Interval Variables	
Default Input Method	Mean
Default Target Method	Distribution
Default Constant Value	
Default Character Value	10
Default Number Value	.
Method Options	
Random Seed	12345
Tuning Parameters	...
Tree Imputation	...

20. ábra Hiányzó adatok pótlásának beállításai

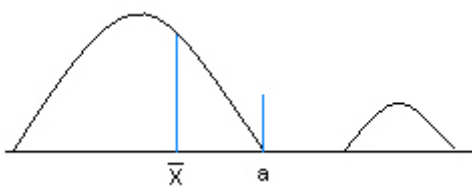
Egy *StatFeltárás* csomópontot 19. ábra alapján bekötve, ellenőrizhetjük a pótlás eredményét.

### 5.3.1. Csere csomópont használata (Replacement)

Ezt a csomópontot a legtöbb esetben az előző *Pótlás* csomóponttal együtt használják. Ez a csomópont a *Módosítás* eszköztárban található.

Funkcióját tekintve a legegyszerűbb magyarázatot a 21. ábra mutatja. Legyen adott egy változónk, mely ilyen eloszlással rendelkezik és tartalmaz hiányzó értékeket. A cél a hiányzó értékek pótlása, de előtte alkalmazni kell a *Csere* csomópontot. Ez azért fontos, mert így az  $a$ -nál nagyobb értékeket nem kell törölnünk, így nem veszítünk egyéb információkat, hanem csak alkalmazzuk a csomópontot és ezeket az értékeket az  $\bar{X}$  átlaggal pótoljuk.

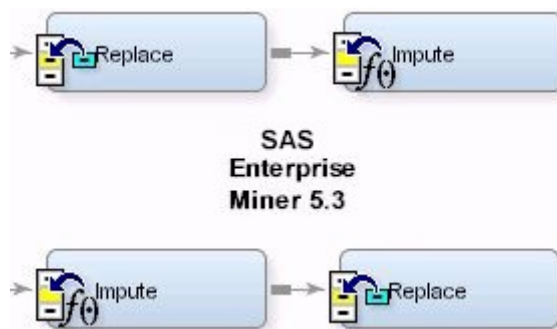
Ezek után már foglalkozhatunk a hiányzó értékek pótlásával, hiszen egy jobban központosult eloszlás áll rendelkezésünkre.



21. ábra Csere csomópont alkalmazása

Természetesen az is megoldható, hogy előbb pótoljuk a hiányzó értékeket, majd ezután alkalmazzuk a *Csere* csomópontot. (A kapott eredmény nem azonos az előzővel)

Az, hogy melyik esetet alkalmazzuk, nagymértékben befolyásolja a változó leíró statisztikai jellemzői.



22. ábra A pótlás és a csere két változata

## 5.4. Új szabály létrehozása

A következőkben bemutatásra kerül, hogyan lehet szabályokat képezni és az eredményeket megjeleníteni. A „kékszalag” adatállományunkat felhasználva alkossuk meg saját szabályunkat.

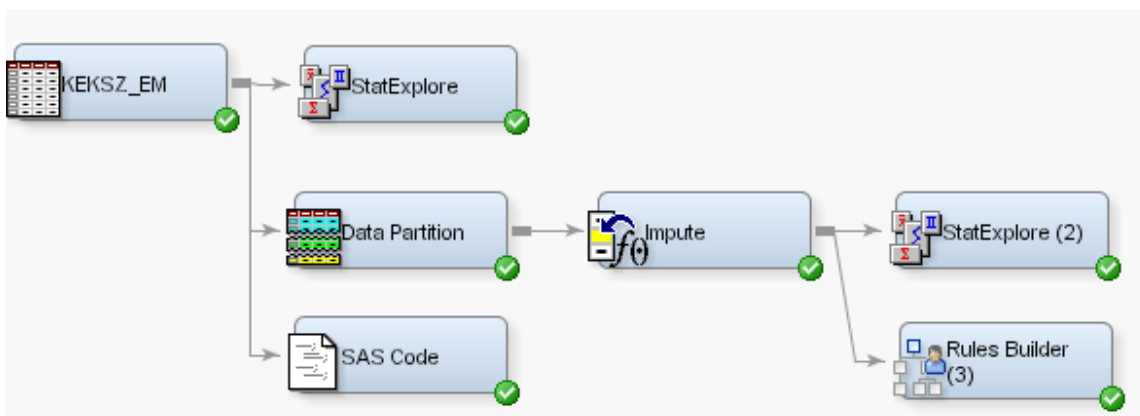
Az előző pontban pótoltuk az összes hiányzó intervallum típusú változót, így ezeket a változókat nyugodtan felhasználhatjuk, hiszen nem tartalmaznak hiányzó értéket és így nem képesek torzítani az eredményt. A következő intervallum típusú változók állnak rendelkezésre:

- Befutási idő
- Futott idő
- Korrigált idő
- YS (Yardstick szám) – minél kisebb egy hajó YS száma, annál gyorsabb

Ezeket az adatokat felhasználva alkossuk meg a következő szabályt.

*A futott idő alapján válogassuk 3 csoportba a hajókat. (gyors, átlagos, lassú) Ezek után bontsuk további két részre a lassú csoportba tartozó hajókat (lassú, nagyon lassú) a YS szám alapján.*

A szabály megalkotásához használjuk a **Módosítás** eszköztár **Szabályalkotó (Rules Builder)** csomópontját. Kössük ezt a csomópontot a 21. ábra szerint.



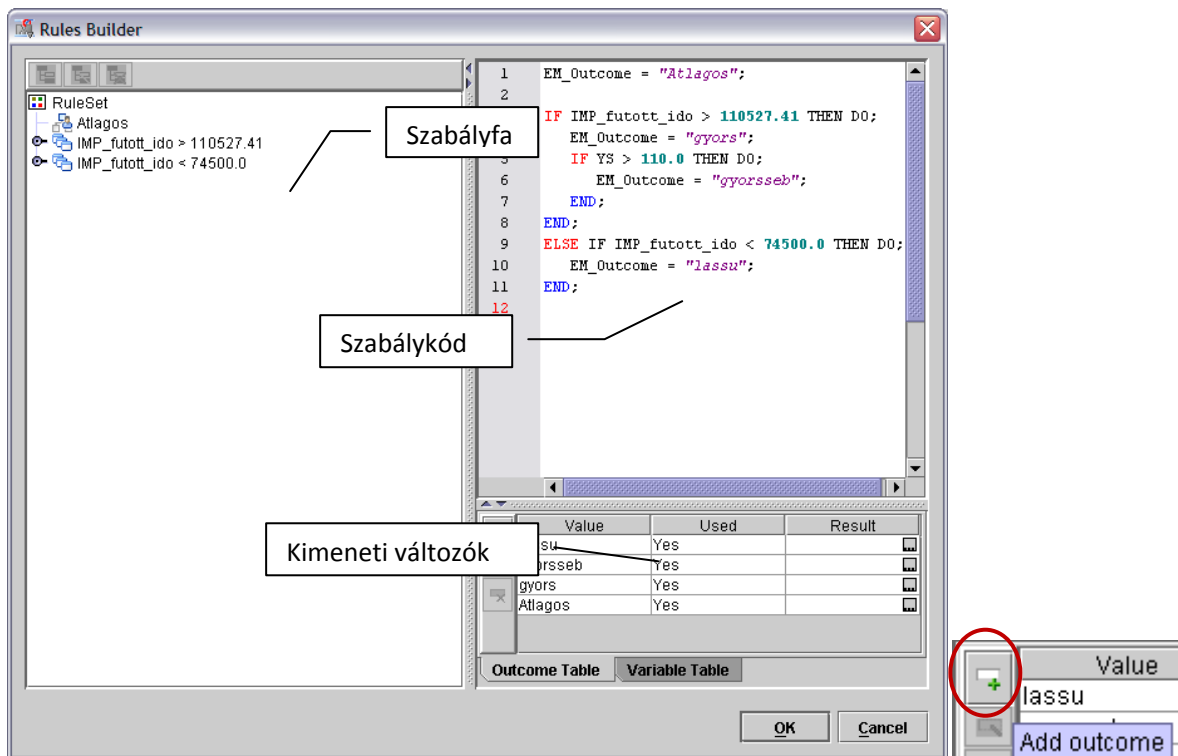
23. ábra Folyamatábra (szabályalkotóval)

**FONTOS!** Mielőtt elkezdenénk szerkeszteni a szabályunkat, legelső lépés, hogy futtassuk a csomópontot az alapbeállításokkal. Jobb klikk a csomóponton, majd Futtatás. (Ekkor egy zöld pipa jelenik meg a csomópont jobb alsó sarkában.)

NOTES		...
<b>Train</b>		
Variables		...
Rules		...
<b>Score</b>		
Output Variable Role	Comment	

24. ábra Szabály megadása

A fenti ábrán látható *Szabály* sor végén található gombra kattintva adhatjuk meg a szabályunkat. Ekkor a *Szabályalkotó* oldal jelenik meg.



25. ábra Szabályalkotó oldal

Kimeneti változókat kell megadnunk. Ezt a 23. ábra jobb oldalán található piros körben lévő gomb segítségével tehetjük meg. Adjuk meg névnek a „lassu”, „gyors”, „leglassabb” eseteket. Majd a „None” nevűt pedig meg kell változtatnunk „Atlagos”-ra. Ezt az előző gomb alatt található másik gombbal adhatjuk meg. (Edit outcome)

A következő lépés a szabályfa megalkotása. (24. ábra)



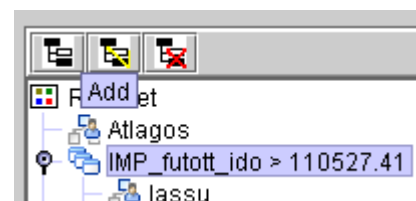
26. ábra Szabályfa alkotás

Kiválasztjuk a futott\_ido változót, majd megadjuk a vágási értéket. (operátor: > érték: 110527) EM\_outcome értéknek pedig adjuk meg a „Lassu” változót.

Következő lépésként ezt meg kell ismételnünk, de más értékkel. (operátor: < érték: 74500) EM\_outcome: „gyors”

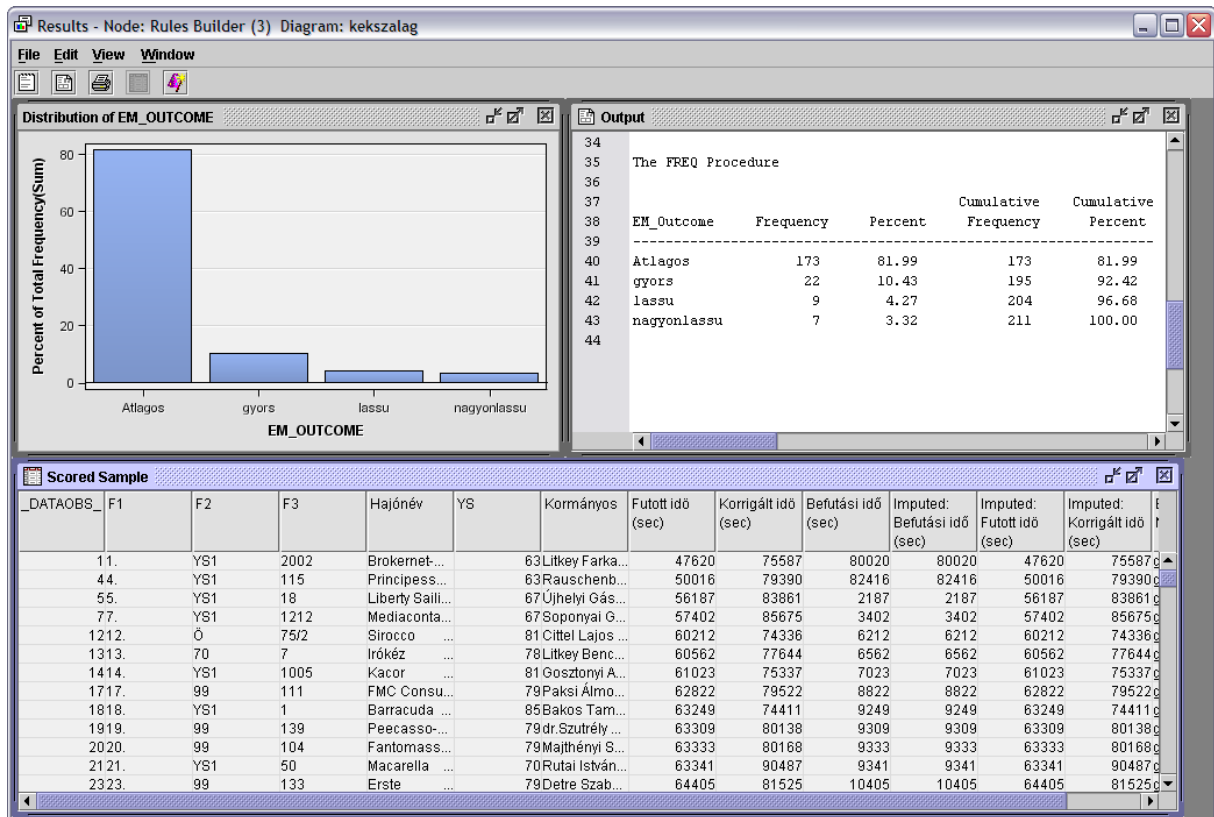
Most el kell végeznünk a lassú szerinti felbontást a YS szám alapján.

Klikk az IMP\_futott\_ido > 110527.41 szabályra, majd (Add) hozzáadás. Ekkor a fentiek alapján:  
Változó: YS operátor: > érték: 110.0  
EM\_outcome: „nagyonlassu”



27. ábra További bontás

Most, hogy megalkottuk a szabályunkat futtassuk ismételten a csomópontot (Jobb klikk / Futtatás), majd tekintsük meg a kapott eredményt. (26. ábra)

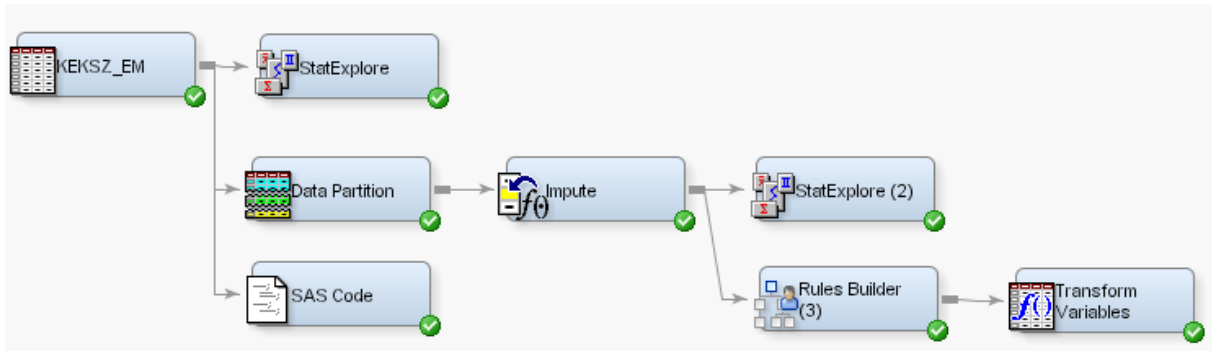


28. ábra A szabály eredménye

## 5.5. Új változó létrehozása

Az adatmódosítás egyik legfontosabb lépése új, képzett változó létrehozása. Új változót a *Változó transzformálás (Transform Variables)* csomópont segítségével tudok létrehozni. Ez a csomópont a *Módosítás* eszköztáron található.

Húzzuk ezt a csomópontot a munkaterületre és kössük a szabályalkotó után. (27. ábra)



29. ábra Folyamatábra

Property	Value
<b>General</b>	
Node ID	Trans
Imported Data	...
Exported Data	...
Notes	...
<b>Train</b>	
Variables	...
Formulas	...
Interactions	...
SAS Code	...
Default Methods	
Interval Inputs	None
Interval Targets	None
Class Inputs	None
Class Targets	None
Treat Missing as Level	No
Sample Properties	
Method	Random
Size	Max
Random Seed	12345
Grouping Method	
Cutoff Value	0.5
Group Missing	No
Missing Values	Use in Results

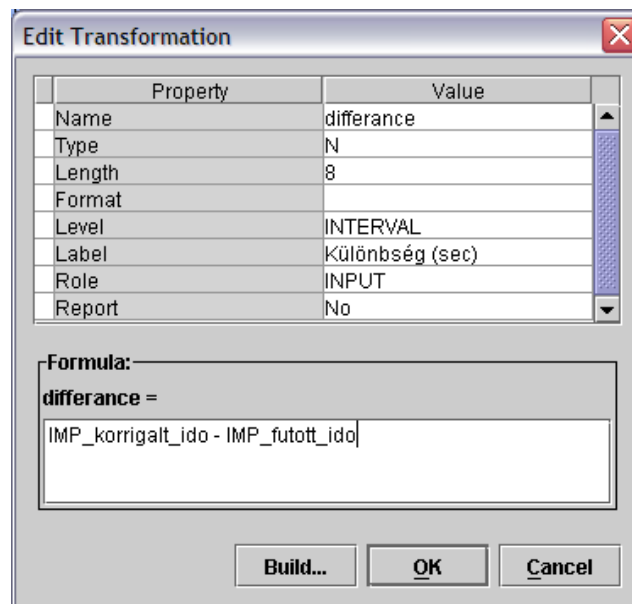
Metódus  
beállítások

30. ábra Beállítási lehetőségek

A beállítások között tekintjük az *Alap metódusok (Default Methods)* részt. Itt lehetőség van intervallum és osztályozó változókhoz generálni a kiválasztott metódusnak megfelelő változót. A metódusok között az alábbiak szerepelnek: négyzet, négyzetgyök, logaritmus, inverz, standardizált... Ha itt kiválasztunk egy metódust az intervallum vagy osztályozó típusú változókra, akkor az összes azonos típusú változóra végrehajtodik.

Mi azonban most egy teljesen új változót kívánunk létrehozni. Ezt a *Formulák (Formulas)* pontban tehetjük meg. Ezután a megjelenő ablak tetején válasszuk a **Létrehoz** gombot. Így az alább látható ablak jelenik meg, amit a következő módon töltünk ki. (29. ábra)

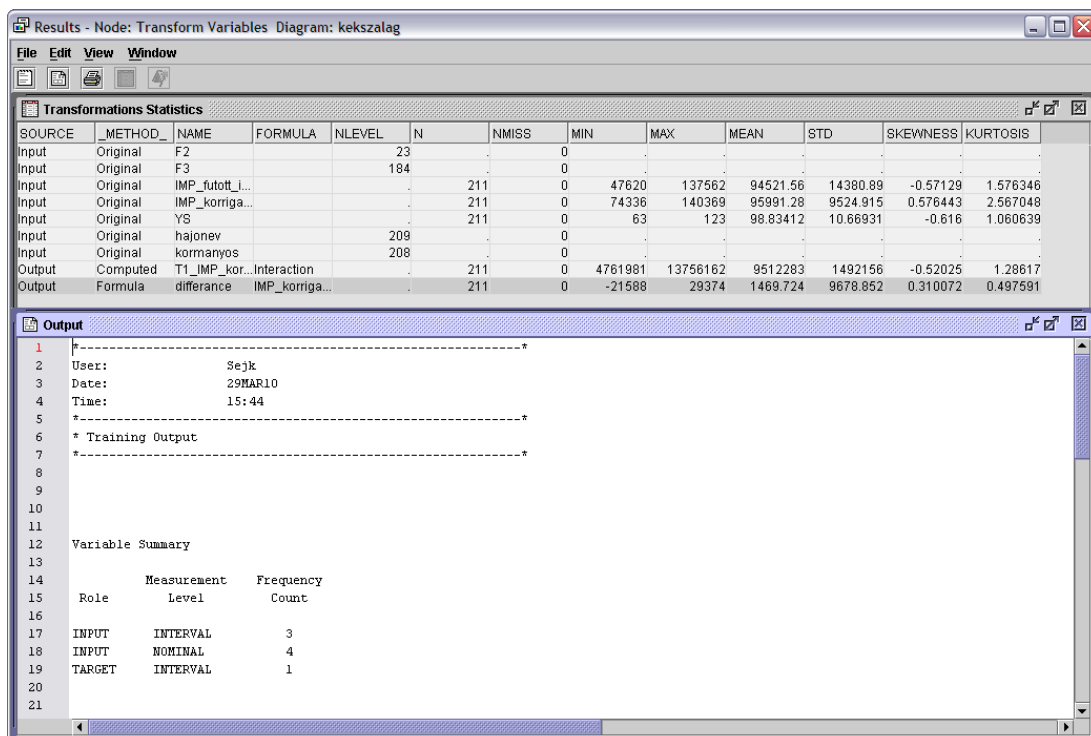
A formula legyen Korrigált idő – futott idő. Ezt az **Épít...** gomb segítségével előugró kifejezés építővel vihetjük be. Itt válasszuk a **Változók listája (Variables List)** fület majd a kívánt változón az **Beillesztés** gombot.



31. ábra Változó létrehozása

Az 5.3-as változat újítása, hogy lehetőség van *Interakció* (Interaction -28. ábra) segítségével változókat létrehozni. Az interakció alatt itt a szorzást kell érteni. Azaz az itt megadott változókból szorzatpárokat képez a Miner és ezeket egy változóba képezi le.

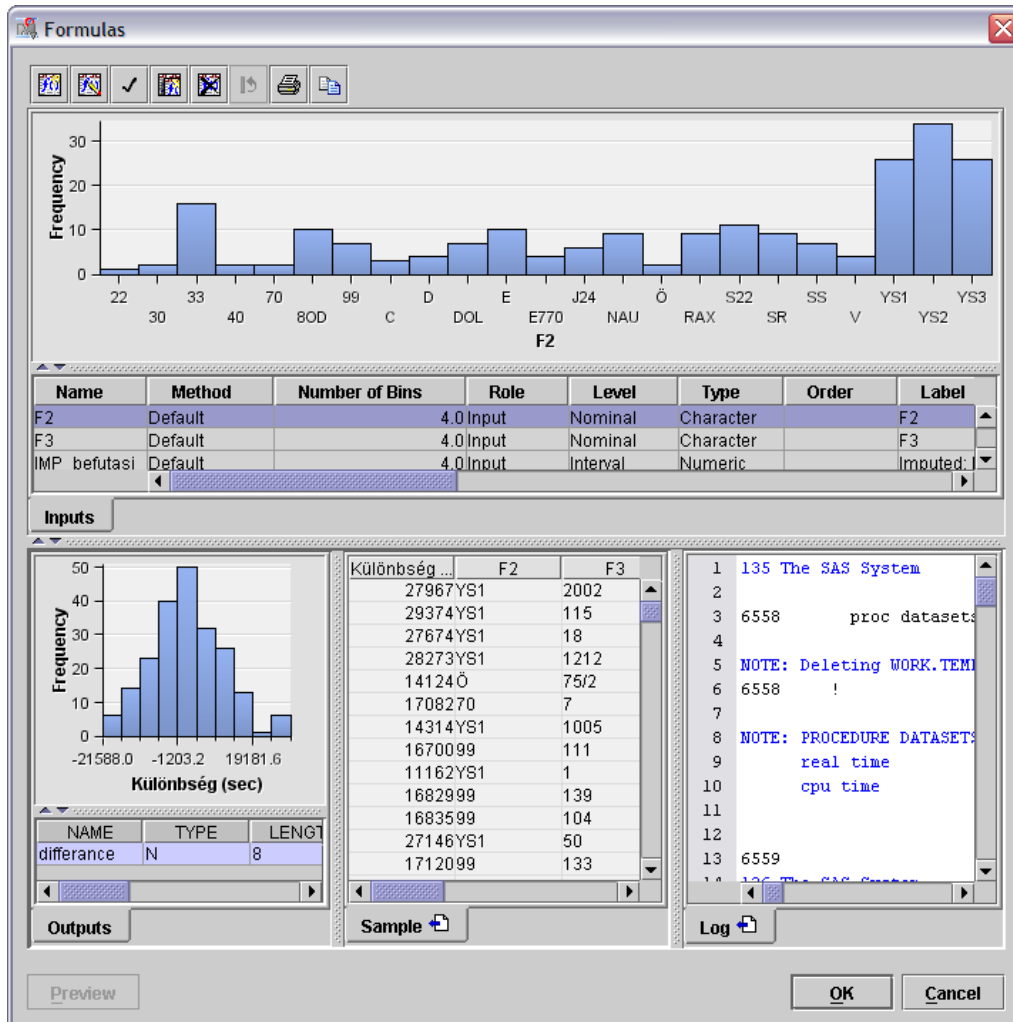
Ezek után futtassuk a csomópontunkat a **Futtatás** paranccsal.



32. ábra Kapott eredmény

A kapott táblázat alján láthatjuk a generált változókat. A formula metódus a fentebb definiált változót jelenti, míg a computed az interakcióval létrehozott változóra utal.

Ha a futás után ismételen kiválasztjuk a beállítások között (28. ábra) a formula utasítást, akkor a kapott ablakban látható lesz a generált változó hisztogramja és mellette táblázatos formában meg is tekinthetjük az eredményeket.

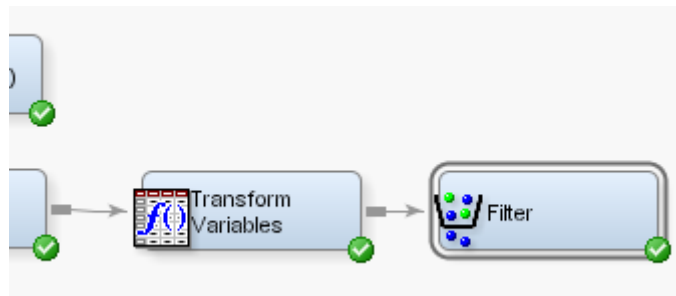


33. ábra A generált változó eloszlása



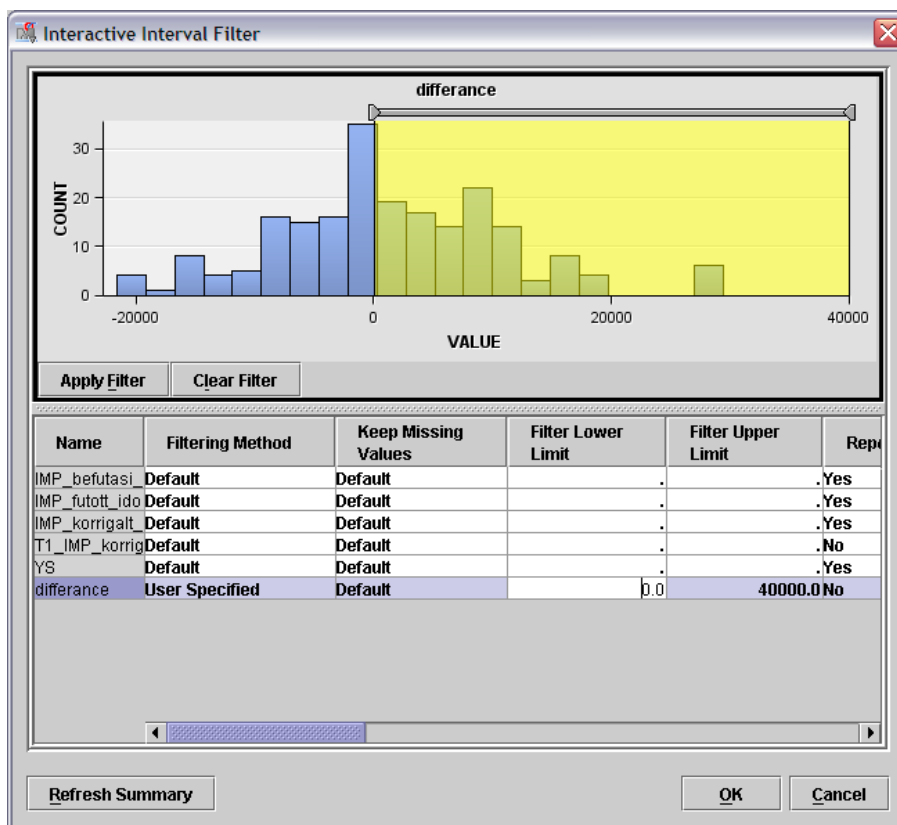
## 5.6. Szűrő (Filter) alkalmazása

Ez a csomópont a *Minta (Sample)* tárban található **Szűrő** néven. Azt szeretnénk elérni, hogy az előző részben létrehozott változó értékei közül kiszűrjük a negatív értékűeket.



34. ábra Szűrő csomópont elhelyezése

A Filter csomópont beállítási lehetőségei nagyon hasonlóak az előző csomópontéhoz. Itt is külön kezelendő az intervallum és az osztályozó típusú változó. Mivel az előző pontban létrehozott változó intervallum típusú, így az „Interval Variables...” sort kell keresni a bal oldali beállítások között. Erre kattintva a következő ablakot kapjuk.



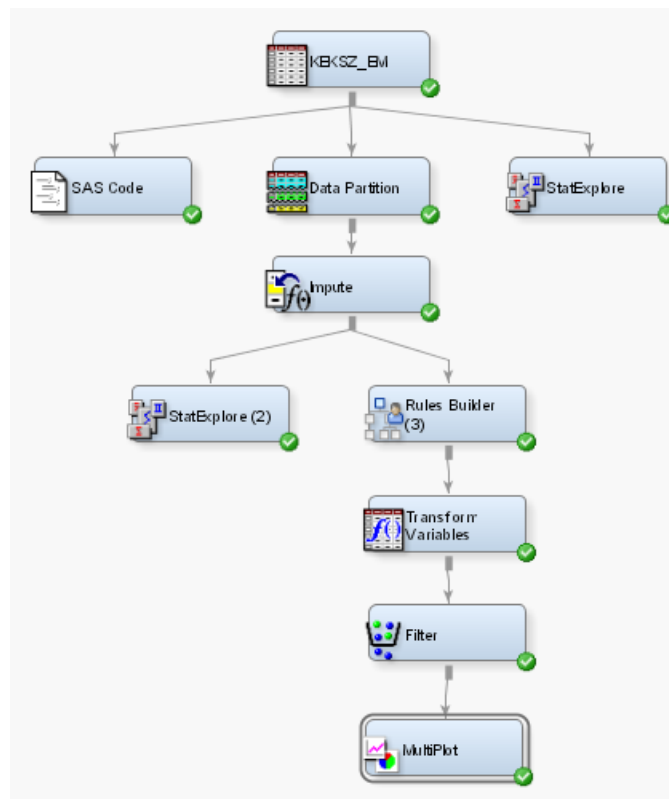
35. ábra Szűrési kritérium

Válasszuk ki a „difference” változót, majd vagy begépeljük a kívánt értéket, vagy a hisztogram felett található csúszka segítségével tehetjük meg. (33. ábra)

Ezek után futtassuk a csomópontunkat.

## 5.7. Multiplot használata

A kapott eredmények áttekintéséhez használjuk a **Multiplot** csomópontot. Ezt a **Feltárás** eszköztárban találjuk. Kössük ezt a csomópontot az előző csomópont végére.



36. ábra A teljes folyamatábra

A Multiplot esetén lehetőség van megadni, hogy milyen grafikont készítsen a Miner. Lehet oszlopdiagram, vagy pontfelhő diagram.

Függőleges vagy vízszintes elrendezésben jelenjen meg.

A célváltozóra lehet kérni statisztikát.

Milyen statisztikát számítson: százalék vagy gyakoriság.

Hiányzó értékek szerepeljenek e a diagramokon.

Property	Value
<b>General</b>	
Node ID	Plot
Imported Data	...
Exported Data	...
Notes	...
<b>Train</b>	
Variables	...
Type of Charts	Bar Charts
<b>Bar Chart Options</b>	
Graph Orientation	Vertical
Include Missing Values	Yes
Interval Target Charts	Mean
Show Values	Yes
Statistic	Freq
Numeric Threshold	20
<b>Scatter Options</b>	
Confidence Interval	Yes
Regression Equation	No
Regression Type	Linear
<b>Status</b>	
Create Time	2010.03.30. 10:38
Run Id	445d7253-3b33-46df-b
Last Error	...

37. ábra Multiplot beállítási lehetőségei

## 5.8. Elrendezés megváltoztatása

Változtassuk meg a nézetet, hogy jobban áttekinthető legyen a folyamatábra. Ehhez klikkeljünk a munkaterületre (Workspace -1ábra Kezelőfelület), majd *Elrendezés/Vertikális (Layout / Vertical)*.