ChatGPT ötletgazdag tanácstalankodók számára

Íme egy példa, amely bemutatja, hogyan lehet az **Excel Solver** segítségével generálni egy adathalmazt, amely megfelel a kívánt statisztikai jellemzőknek, például átlagnak, szórásnak, minimum- és maximumértékeknek, valamint egy adott korrelációs mátrixnak. *(szerk.: ez egyben a prompt is lehetett volna:-)*

**Feladat: Valósághű adathalmaz generálása adott statisztikai jellemzők alapján**

**Adott statisztikai jellemzők:**

* Átlag (Mean)
* Szórás (Standard deviation)
* Minimum és Maximum értékek
* Korrelációs mátrix

**1. Lépés: Minta adatok és követelmények beállítása**

Tegyük fel, hogy két változóval dolgozunk (pl. X és Y), és az alábbi céljellemzőket szeretnénk:

| **Mérték** | **X** | **Y** |
| --- | --- | --- |
| Átlag | 50 | 30 |
| Szórás | 10 | 5 |
| Min | 30 | 20 |
| Max | 70 | 40 |

A kívánt korrelációs együttható: **0.8** (X és Y között).

**Excel-beállítások:**

1. Hozz létre egy táblázatot az X és Y változók számára (pl. 20 sorban), majd töltsd fel őket véletlenszámokkal (pl. =RANDBETWEEN(30,70) az X oszlopban).
2. Számold ki az aktuális statisztikai jellemzőket:
	* **Átlag:** =AVERAGE(X:X)
	* **Szórás:** =STDEV.P(X:X)
	* **Minimum:** =MIN(X:X)
	* **Maximum:** =MAX(X:X)
	* **Korreláció:** =CORREL(X:X, Y:Y)

**2. Lépés: Solver konfigurálása**

1. Nyisd meg az **"Adatok" (Data)** fülön az **"Solver"** eszközt.
*(Ha nincs telepítve, aktiváld az Excel bővítmények között.)*
2. Állítsd be a következőket a Solverben:
	* **Célcella (Set Objective):**
		+ Állítsd a korreláció cellára (például ahol =CORREL(X:X, Y:Y) van).
		+ Válaszd a **Minimization** (vagy értékre állítás: 0.8) opciót.
	* **Döntési változók (By Changing Variable Cells):**
		+ Az X és Y adatok cellatartománya (pl. A2:B21).
	* **Korlátozások (Constraints):**
		+ Átlag: AVERAGE(X:X) = 50, AVERAGE(Y:Y) = 30
		+ Szórás: STDEV.P(X:X) = 10, STDEV.P(Y:Y) = 5
		+ Minimum/Maximum: pl. MIN(X:X) >= 30, MAX(X:X) <= 70
		+ Korreláció: CORREL(X:X, Y:Y) = 0.8
3. Válaszd az **"Evolúciós"** vagy **"GRG Nonlinear"** algoritmust, mivel a korrelációs optimalizálás nemlineáris probléma.
4. Kattints a **"Megoldás" (Solve)** gombra.

**3. Lépés: Az eredmények elemzése**

* Ellenőrizd, hogy az adatok megfelelnek-e az elvárásoknak.
* Ha nem sikerült pontosan elérni a célokat, finomítsd a korlátokat vagy próbáld újra az evolúciós algoritmussal.

Ha szeretnéd, segíthetek egy részletesebb vizuális útmutató vagy egy példafájl elkészítésében.

S erre jutott az önállóan megoldást nem találó ember:

Az alábbi táblázat bemutatja az Excel Solver eszköz felhasználását egy valósághű adatvagyon létrehozására. A feladat során az X és Y adatsorok statisztikai jellemzőit állítjuk be és optimalizáljuk?, hogy az előre meghatározott átlag, szórás, minimum és maximum értékeknek megfeleljenek, valamint egy adott korrelációs összefüggést tartsanak fenn.

Először egy Excel táblázatot hozzunk létre, amely tartalmazza az X és Y adatsorokat. Az X adatok 30 és 70 között, míg az Y adatok 20 és 40 közötti véletlenszámokat tartalmaznak. Ezek az adatok kezdetben véletlenszerűen kerülnek generálásra, majd a Solver segítségével optimalizáljuk? őket.

A Solver eszközt az Excel "Adatok" fülön találjuk meg. Ha nem lenne elérhető, aktiválni kell az Excel bővítmények menüpontja alatt.

A Solver konfigurálása során először meg kell határoznunk a célcella („Set Objective”), amely az =CORREL(X:X, Y:Y) függvény által kiszámolt korrelációt tartalmazza. Célként az 0.8? értékre történő beállítást választjuk, hogy az adatsorok erős pozitív kapcsolatot mutassanak.

A döntési változók („By Changing Variable Cells”) azok az Excel cellák lesznek, amelyek az X és Y adatok értékeit tartalmazzák. Ezek az A2:B21? tartományba esnek. Ezen cellák értékeit a Solver az optimalizáció során módosítja, hogy megfeleljenek a megadott követelményeknek.

A korlátozások („Constraints”) során a következő feltételeket adjuk meg:

* Az X adatok átlaga legyen 50, az Y adatok átlaga 30. (a mindenkori generált adatokra vonatkozó átlag/szórás/max/min/korreláció eltérése legyen a célátlagtól minimális egy nagy hibafüggvény részeként?)
* Az X adatok szórása 10, az Y adatok szórása 5. (a mindenkori generált adatokra vonatkozó átlag/szórás/max/min/korreláció eltérése legyen a célátlagtól minimális egy nagy hibafüggvény részeként?)
* Az X minimuma legalább 30, maximuma legfeljebb 70. (a mindenkori generált adatokra vonatkozó átlag/szórás/max/min/korreláció eltérése legyen a célátlagtól minimális egy nagy hibafüggvény részeként?)
* Az Y minimuma legalább 20, maximuma legfeljebb 40. (a mindenkori generált adatokra vonatkozó átlag/szórás/max/min/korreláció eltérése legyen a célátlagtól minimális egy nagy hibafüggvény részeként?)
* A korreláció az X és Y adatok között legyen pontosan? 0.8. (a mindenkori generált adatokra vonatkozó átlag/szórás/max/min/korreláció eltérése legyen a célátlagtól minimális egy nagy hibafüggvény részeként?)

A megoldáshoz az "Evolúcios" vagy a "GRG Nonlinear" algoritmust kell használni, mivel a korreláció optimalizálása nemlineáris problémának minősül.

Miután a Solver lefut, az eredmények kiértékelése során ellenőrizni kell az átlag, szórás és korreláció értékeit, hogy azok megfelelnek-e a megadott követelményeknek. (a mindenkori generált adatokra vonatkozó átlag/szórás/max/min/korreláció eltérése legyen a célátlagtól minimális egy nagy hibafüggvény részeként?)

A korrelációs mátrix (jelenleg 2\*2-es, aminek az átlója 1 és 1, így a maradék két cella +/- korreláció az X és Y között) elemzése során figyelni kell arra, hogy az eredeti adatok szerinti korreláció (általában a Pearson-féle?) mennyiben tér el a várt értéktől. Ha a Solver megfelelően dolgozott, az elért korreláció közel lesz az elvárt 0.8 értékhez, így biztosítva, hogy az X és Y változók között az előírt kapcsolat fennáll.

A Solver futtatása után az eredményeket rögzíteni kell, és ellenőrizni, hogy az optimalizált értékek megfelelnek-e az elvárásoknak.



XLSX-melléklet hamarosan

Anomáliák

A sárga korrelációs cellákból csak egy lehet és ez a zöld korrelációs cella értékével közel azonos illene, hogy legyen, lévén a korrelációszámításhoz két változó kell…

A vizualizáció kapcsán nem egységes a tizedes jegyek száma, ami nem tartalmi hiba, de csúnya és nehezen értelmezhetővé teszi a számok nagyságrendjét = gondatlan, az Olvasóra nem gondoló megoldás…

Az XLS-ábráról hiányzik a Solver átlal minimalizálandó hiba, mint olyan…

Sőt, a kékes hátterű cellák szerepe is kérdéses?!

A minél közelebb a sokkomponensű célhoz annál jobb elvárást többféleképpen lehet közelíteni, de pl. az azonosságok kikényszerítése antagonizmusokhoz vezethet…

A statisztikai mutatók idealizált értékeinek FORRÁSA: a szakirodalom, ahol azonban eltérő adathalmazok jellemzőire vonatkozó mutatószámok (átlag, szórás, max, min, korreláció) lelhetők fel, mely önmagában is lehet antagonizmusok (solver-futás-akadályok) megalapozója…