**SEM elemzés**

# **2023.05.11. Statisztikai terv**

## 1. Faktorok azonosítása

A változókat/teszteket három csoportba osztottam egy előzetes táblázat alapján: nyelvi tesztek + ToM, background tesztek, SL tesztek. A csoportokon belül feltáró faktorelemzéssel kerestem olyan tesztcsoportokat, amely az eredmények alapján többnyire egy dolgot mérnek, azaz a pontszámaik egymással nagyban korrelálnak. A nyelvi tesztek közül egy ilyen csoport van, a háttérképességeknél három:

|  |  |  |
| --- | --- | --- |
| variable\_cluster | variable\_name | variable\_type |
|  |
| OFFLINE LANG | rec\_voc, trog\_prag, menyet | language |  |
| RTM+S | rec\_voc, trog\_prag, menyet, SAT | language, ToM |  |
| CC | n\_back1, n\_back2, n\_back3 | background |  |
| WM/STM | digit\_span\_forward, digit\_span\_backward | background |  |
| VISual PS | stroop\_RT, proc\_speed\_vis\_RT, proc\_speed\_vis\_de\_RT | background |  |

Az SL-tesztek közül a vizuális szegmentálásos teszteket kivettem, mert túl kevés adat társul hozzájuk, a nyelvi szegmentálásos tesztek közül a reakcióidőt mérő tesztek alkotnak egy csoportot (SEGM\_AL\_medRT). A produkciós változó ezen kívül önállóan is jó mutatónak tűnik. Az accuracy változók státusza bizonytalan, lehet, hogy ki lehet hozni még egy tesztcsoportot, de egyelőre nem sikerült.

## 2. Változók közötti összefüggések

Az azonosított faktorokat, illetve a kimaradt tételeket prediktor változókként lineáris regressziós modellbe tettem, ahol a kimeneti változót az SL-változók jelentették (SEGM\_AL\_medRT-csoport, SEGM\_AL\_prod). A csatolt táblázatban a cellában szereplő érték a regressziós koefficienst jelöli, zölddel jeleztem a szignifikáns összefüggéseket, narancssárgával a nem szignifikánsakat. A halványzöld azt jelzi, hogy ugyan van szignifikáns kapcsolat a két változó(csoport) között, a kapcsolat mértéke kicsi.

Ami az eredmények alapján érdekes, hogy a SEGM\_produkciós változót jól bejósolja az azonosított nyelvi faktor, illetve közepesen erősen az n\_back és a digit\_span faktor, illetve önálló változóként a proc\_speed\_ac\_dec\_RT.

A nyelvi és háttérképességek faktorai között magas együttjárás tapasztalható, érdemes tehát még ezek összefüggéseinek vizsgálata is.

## 3. Hipotézistesztelés

A modell építésében a következő lépést a lehetséges mediátorok vizsgálata jelenti. Ehhez viszont szükséges előzetes hipotéziseket alkotni (hogy mi mi által hat), mert ezt nem lehet adatvezérelten végezni.

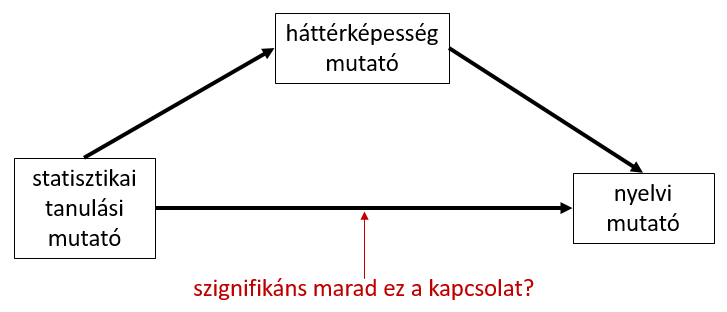
### *3.1. Kutatási kérdések:*

I. *Kapcsolatban vannak egymással a statisztikai tanulási és a nyelvi képességek?* (Itt nem feltétlenül feltételezünk irányt – alapvetően az a hipotézis, hogy a statisztikai tanulási képességek hatással vannak a nyelvi képességekre, de lehet visszafelé hatás is.)

* Itt minden nyelvi mutató és SL mutató esetén megnézzük, hogy szignifikánsan kapcsolatban állnak-e egymással.

II. *Ha igen, ezt a kapcsolatot mennyiben mediálják a nemnyelvi háttérképességek?* (Eredetileg processing speed, rövid távú memória és kognitív kontroll (gátlás) lett volna, de más faktorokat találtunk.) Ennek a kérdésnek a vizsgálata két lépésből áll:

* Először minden háttérképesség esetén megnézzük, hogy szignifikánsan kapcsolatban áll-e a nyelvi mutatókkal
* Utána minden háttérképesség esetén megnézzük, hogy szignifikánsan kapcsolatban áll-e az SL mutatókkal
* Utána azokban az esetekben, amikor 1) szignifikáns a kapcsolat a nyelvi mutató és az SL mutató között, 2) szignifikáns a kapcsolat a háttérmutató és az SL mutató között, és 3) a háttérmutató és a nyelvi mutató közötti kapcsolat is szignifikáns, akkor megnézzük, hogy a háttérképesség mediálja-e a SL-nyelvi kapcsolatot:



* Persze lehet, hogy erre van olyan módszer/modell, amelyben lehet egyszerre az összes kérdést vizsgálni, és nem kell egymás utáni regressziós elemzéseket futtatni.

III. *Bónusz kérdés: mi a helyzet, ha bevesszük a vizuális szegmentálást is?*

# **2023.06.07.**

## 1.

…

## 2.

…

## 3.

…

## X. További kérdések a későbbiekre

* Amiatt gyengült ebben az esetben az offline SL – offline nyelv közötti kapcsolat, mert sok más kapcsolatot bevettünk a nagy modellbe?
* Lehetséges egy kétirányúság a nyelvi statisztikai tanulási feladat és a nyelvi feladatok között? Ha igen, ez hogyan írható le a modellben?
* Más változók esetében kell feltételeznünk kétirányúságot?