



Statisztika

Bevezetés

- Mire jó a statisztika?
- Számszerű adatok leírása
- Adatok értelmezése
- Összefüggések keresése adatok, csoportok, változók között
- Statisztikai tesztek



Idézzük fel...

- **Kvalitatív kutatás**
- **Kvantitatív kutatás**
- **Kísérleti kutatás**
- **Nem kísérleti kutatás**
- **Függő változó**
- **Független változó**



Statisztikai próbák

- Megvizsgáljuk, hogy a csoportok közti eltérés a véletlennek köszönhető-e
- A csoportok közti különbség mekkora eséllyel köszönhető a független változónak?
- Szignifikancia: matematikai értelemben véve
- Általában $p < 0,05$ -> szignifikáns (5%-nál kisebb az esélye, hogy a véletlennek köszönhetőek az eredmények)



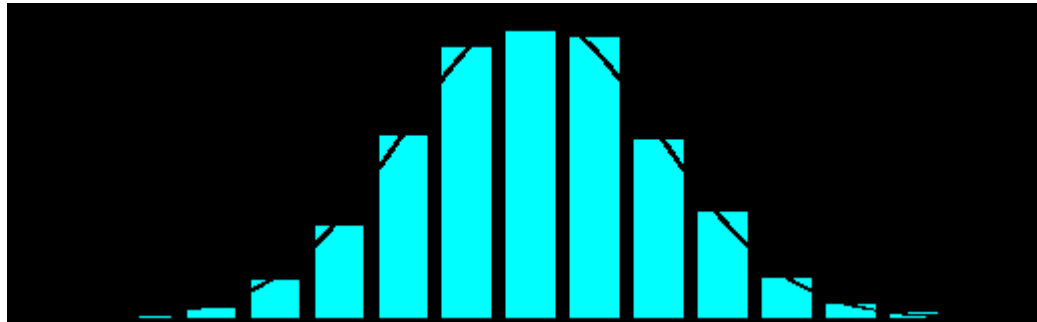
Vizsgált problémák

- Alzheimer-kórosok és egészséges kontrollszemélyek hezitációinak száma spontán beszédben
- A szintaktikai szerkezet hatása óvodás gyerekek szótanulására
- Függőségi relációk eloszlása a Szeged Dependencia Treebankben
- FX-ek típusainak eloszlása a Szeged Treebankben
- Függőségi relációk eloszlása a névszók esetében



t-próba

- A használt adatok intervallum skálán helyezkednek el (x és y között levő számok)
- Normális az értékek eloszlása (vannak magas és alacsony pontszámok is, de a legtöbben a középérték környékén teljesítenek) – haranggörbe
- $20 <$ a résztvevők száma
- Két csoportunk van, egy független változóval



t-próba

- **Egymintás t-próba:** ugyanazok a személyek vesznek részt két különböző kísérleti fázisban (pl. vezetés józanul és ittasan)
- **Kétmintás t-próba:** eltérő csoportok vesznek részt a különböző fázisokban (pl. nyelvtanítási módszerek összevetése az a) és b) osztályokban)
- **Példa:** alzheimeresek - stat_korpnyelv.xlsx



Khi² (χ^2)-próba

- Nominális adatok esetén alkalmazható (számszerűen nem írható le)
- Pl. kérdőíves felmérések adatainak elemzése
- Nők vagy férfiak szavaznak inkább az adott pártra?
- Iskolai végzettség befolyásolja-e a fizetés nagyságát?
- Gyakorisági adatokat nézünk (nem százalékokat)



Khi² (χ^2)-próba

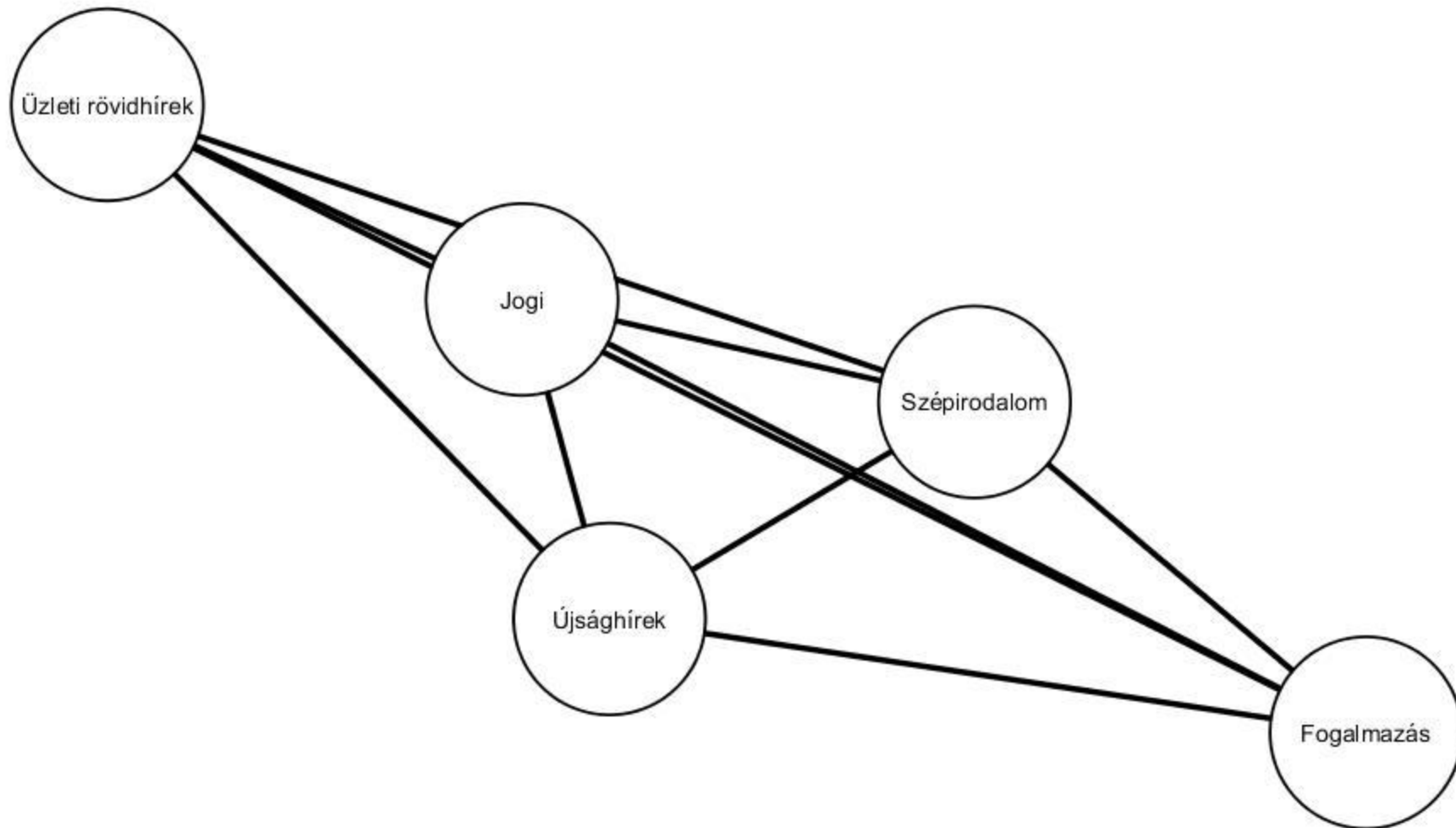
- **Megfigyelt gyakoriság:** a tényleges adatok
- **Várható gyakoriság:** mit kapnánk, ha nem lenne összefüggés (nem befolyásolná a független változó a függő változót)?
- **Példa:** óvodások - stat_korpnyelv.xlsx



Kendall's W

- Gyakorisági rangsorokkal dolgozik
- Adatok gyakorisága
- Szövegtípusokban milyen a szófaji eloszlás?
- Hasonlósági mérték:
 - 0 és 1 között
 - Minél nagyobb, annál hasonlóbbak
- χ^2 -próbára vezethető vissza a szignifikancia kiszámítása
- Példa: FX-ek, függőségi relációk - stat_korpanyelv.xlsx





ANOVA

- **Varianciaanalízis**
- **Kettőnél több csoport**
- **Egynél több független változó**
- **Mindkettő**
- **Két független változó (pl. nem és szak):
kétirányú ANOVA**
- **Példa: függőségi relációk és névszók -
stat_korpanyelv.xlsx**

