

# Kvantitatív Alapok

Kovács Márton

8/27/23

## **Table of contents**

# Előszó

Ez a könyv a Neumann János Egyetem és az MNB Intézet által indított Nemzetközi Gazdaság és Gazdálkodás mesterszakának *Kvantitatív alapok* nevű tantárgyához készült. Céлом a hallgatók felkészülését segíteni azzal, hogy egy helyen, átlátható formában közlöm a tantárgyhoz készült jegyzeteket, példákat, gyakorlatokat és ábrákat. A Kvantitatív alapok tantárgy felépítésénél akárcsak ezen könyv megírásánál nagyban merítettem Russ Poldrack *Statistical Thinking for the 21st Century* című munkájából, Nagy Tamás *Adatelemzés és statisztikai következtetés* nevű órai jegyzeteiből és Andy Field *statisztika* tankönyveiből. Ha az olvasó ezen művekből ismerős gondolatokat, példákat vagy magyarázatokat lát, akkor azok nagy valószínűséggel onnan is lettek átvéve. Én magam is nagyrészt tőlük tanultam azt, amit a kvantitatív elemzésről tudok. Ez a három ember nálam sokkal többet gondolkozott azon, hogyan kell hatékony módon statisztikát tanítani egyetemista hallgatóknak, így nem láttam értelmét, hogy a tantárgy felépítését a nulláról kezdjem. Ezúton is szeretném megköszönni a segítségük!

# Köszönetnyilvánítás

Szeretnék továbbá köszönetet mondani Ignits Györgyinek, Kapitány Baláznak, Szászi Barnabásnak és Szécsi Péternek, akik példa adatok megosztásán keresztül segítették a könyv létrejöttét.

A KÖNYVÖN MOST IS AKTÍVAN DOLGOZOM, EZ NEM A VÉGLEGES VÁLTOZAT!

# 1 Bevezetés

A kvantitatív elemzés nem az adatoknál kezdődik. Az eddigi oktatói tapasztalataim alapján amikor egy hallgató elakad egy önálló elemzési feladat során az nem azért van mert helytelenül magolta be az éppen alkalmazott statisztikai teszt előfeltételeit, vagy mert nem tudja a standard hiba képletét. Ezekre a kérdésekre egy gyors Google kereséssel könnyen meg lehet találni a választ.

- Hanem mert nem látja át az empirikus tudományos módszer egészét, annak lépéseit és összefüggéseit
  - Nincs jól megfogalmazott hipotézis
  - Nem feltett kutatási kérdés megválaszolásához szükséges mérőeszközöket használtak
  - Nem látják át a rosszul strukturált adattáblát
  - Nem értik a frekventista hipotézis tesztelés működését
  - És csak ezután jön, hogy nem értik milyen statisztikai eljárást kell alkalmazni és azt hogyan
  - Ezzel a lépéssel kapcsolatban én is gyakran teszek fel kérdéseket kollégáknak és az internetnek (és bíztatok mindenkit, hogy tegyen ugyanígy)

## 2 Tudományos módszer

### 2.1 A tudomány célja

- A tudomány célja, hogy általános magyarázatokkal szolgáljon a minket körülvevő világról
- A tudományos magyarázatok leegyszerűsített formában **írják le** a valóságot
- A tudományos magyarázatok segítségével így képesek vagyunk
  - jól informált **döntéseket** hozni
  - **predikciókat** alkotni új helyzetekre a már meglévő tudásunk alapján
- Később majd látni fogjuk, hogy a kutatási folyamat sok pontján félrecsúszhat az, hogy a kutatásunk alapján megbízható döntéseket hozzunk vagy predikciókat alkossunk
- A tudományos megközelítés azonban nem az egyetlen mód, amelyen keresztül a valóságot magyarázzuk
- Miben más, mint egyéb megismerési módok?
  - Pl: vallás -> autoritás alapú
  - Kritikai megközelítés
  - Tesztelhető (falszifikálható) hipotézisek

### 2.2 Az empirikus tudományos folyamat lépései

1. Általános kérdésfelvetés
  - Ez talán a legegyszerűbb lépés, az ember kíváncsi természetű, szeretné érteni az őt körülvevő valóságot
2. Elmélet alkotás
  - Ahhoz, hogy egy kérdést meg tudjunk válaszolni empirikus módon egy elméletet kell alkotnunk a kérdés mögött álló lehetséges hatásmechanizmusról

- Az elmélet falszifikálható kell legyen
  - Tudunk olyan megfigyeléseket tenni amelyek fényében az elméletet elvetjük
- A jó elméletek
  - Elérés (reach): minél több jelenséget magyaráz meg egy elmélet annál jobb
  - Parszímónia: két versenyző elmélet közül az a jobb, ami kevesebb elemből áll
    - \* Lásd Occam borotvája
  - Konzervatívizmus: Mennyire illeszkedik már meglévő tesztelt és elfogadott elméletekhez

### 3. Hipotézis alkotás

- A hipotézis az elméletünknek egy konkrétan tesztelhető operacionalizált része
- Egy elmélet alapján többféle hipotézist fel lehet állítani
- A hipotézis alapján predikciókat alkotunk, amelyeket tesztelünk
- Egy hipotézis bizonyításával nem bizonyítjuk az elméletet is!
- De tudjuk a teszt eredményei által tovább finomítani az elméletet
- Vagy elvetjük azt
- Fontos, hogy a hipotézis valóban illeszkedjen az adott elmülethez

### 4. Kutatási elrendezés

- A hipotézis tesztelésére annak feltételezései alapján egy kutatást tervezünk
- A kutatás megtervezésénél számos különböző módszertan közül tudunk választani
- Mindegyiknek vannak előnyei és hátránya, a hipotézis és az elérhető erőforrások fényében választunk közülük
  - Kérdőíves
  - Megfigyeléses
  - Kísérleti
- Van pár általános szempont, amelyeket módszertantól függetlenül figyelembe kell vennünk és amelyek befolyásolják az eredményekből levonható következtetéseket
  - Minta felépítése és mintaméret
  - Mérőeszköz és annak pontossága, megbízhatósága, és granularitása

- Ahhoz hogy oksági viszonyt feltételezzünk két jelenség közt nem felelkezhetünk meg az adatgyűjtés módjáról
- Megfigyeléses vizsgálatoknál nem tudjuk biztosan kijelenteni hogy két változó között oksági viszony van
  - Lehetséges hogy egyéb faktorok befolyásolják a kapott együttjárást
  - <https://www.tylervigen.com/spurious-correlations>
- Ezért a kutatók általában az oksági viszony meghatározásához kísérleti kutatási elrendezést használnak (pl: randomized control trial)
  - A lehetséges, de nem vizsgált befolyásoló faktorokat random mintavételezés útján zárják ki az adatgyűjtés során
- Vannak statisztikai módszerek amivel utólag is kontrollálni tudjuk ezeket a változókat, de sokkal nehezebb ezt utólag megtenni
- Ez is kiemeli a jól megtervezett kutatások és részletesen kidolgozott elméletek fontosságát!
- Mintavételezés
  - Az esetek kis részében tudjuk az egész populációt vizsgálni
  - Minták tesztelésén keresztül vonunk le általános következtetéseket
  - De mekkora mintára van szükség ahhoz, hogy a populációra tudjunk következtetni?

## 5. Adatok feldolgozása

- Általában nem szokták külön lépésként kezelni, de az óra témaja és az eredményekre gyakorolt lehetséges hatása miatt fontos megemlíteni
- A kapott adatokat befolyásolja az általunk választott kutatási elrendezés és az adatgyűjtés során fellépő váratlan torzító tényezők
- Nagyon ritka esetben készek az adatgyűjtés után az adatok az elemzésre
- Általában először adat rendezést és adattisztítást kell elvégeznünk

## 6. Statisztikai elemzés

- Az elemzés során használhatunk
  - Leíró statisztikai eljárásokat
  - Következtetési statisztikai eljárásokat



- Rengeteg statisztikai eljárás van, a statisztikusok most is dolgoznak újak kifejlesztésén és validációján
- Az, hogy milyen eljárást választunk függ a hipotézisünktől, a kutatási elrendezésünktől, és az adatgyűjtés során kapott adatoktól
- Többféle lehetséges valid elemzési út is létezik
  - Lást: multi-analyst study

## 7. Eredmények értelmezése

- A statisztikai elemzés eredményeinek értelmezésére a tudományos folyamat összes eddigi lépése kihat
- Kutatóként szeretnénk egyértelmű válaszokat kapni az általunk feltett kérdésekre
  - Sajnos nagyon kevés esetben ez a helyzet
  - Ez részben a tudományos folyamat összetettségének az eredménye
  - Részben ennek a váagnak is szerepe van a p-értékek abuzálásában
  - Egy kutatásból nagyon ritkán kapunk egyértelmű válaszokat
    - \* Lásd: metaanalízis, szisztematikus összefoglaló (systematic review)
- Pár példa helyzet arra, amikor a tudományos módszer összetettsége árnyalja a statisztikai elemzés eredményeinek értelmezését:
  - A választott mérőeszközök valójában nem a vizsgált konstruktumot méri
  - A minta milyensége vagy mérete nem indokolja az eredmények általánosítását
  - Az adattisztítás során olyan megfigyelések is bent maradtak, amelyek szisztematikusan torzítják az eredményeket
- Az értelmezést korlátozó tényezőkről általában a tudományos publikáció limitációk szakaszában számolunk be

## 2.3 A statisztika szerepe a tudományban

- A komplex valóság egyszerűsített leírása, ami egyúttal azt is elmondja mennyire lehetünk bizonytalanok ebben a tudásban
- Miért van szükség a statisztikára?
  - Heurisztikák veszélye
    - \* Példa: USA bűnelkövetések gyakorisága

- Sokkal gyakoribbnak gondoljuk a bűnelkövetések számát, mint amilyen gyakoriak a valóságban
- Elérhetőségi torzítás (kognitív torzítás egyik fajtája)
- Média szerepe?
- Statisztika segít abban hogy tanuljunk az adatokból
  - \* Hogyan frissítsük előzetes tudásunk az új adatok fényében
- A statisztika az adatösszesítésről szól
  - Hogyan tudunk úgy leegyszerűsíteni befogadhatatlan mennyiségű megfigyelést, hogy azok segítségével hipotézisünk tesztelni tudjuk, de fontos információ ne vesszen el
- A statisztika a bizonytalanság kezeléséről szól
  - A világ összetett
  - A legtöbb összetett jelenséget nehéz determinisztikus módon megmagyarázni
  - Bár tudunk ok-okozati kapcsolatot felállítani bizonyos összefüggések között általában ezek nem determinisztikus, hanem probabilisztikus predikciókhoz vezetnek
    - \* Lásd: dohányzás és tüdőrák
      - Tudjuk, hogy a dohányzás növeli a tüdőrák esélyét, azonban nem lehet biztosra mondani, hogyha valaki dohányzik, akkor mindenképp tüdőrákos is lesz
  - A statisztika nem tud bizonyítani (úgy ahogy a matematika tud)
    - \* Hanem evidenciát tud felmutatni egy hipotézis mellett az eddigi megfigyelések fényében, de a bizonytalanságot nem tudjuk nullára csökkenteni

A statisztika a kompromisszumokról szól

- Nincs egyetlen objektív módszer a statisztikában
- Mindig a vizsgált kérdés, az adatgyűjtést és elemzést korlátozó tényezők függvényében kell megtalálnunk a legmeggyőzőbb elemzési utat
  - Ahhoz, hogy a választott elemzési út ténylegesen meggyőző legyen transzparensen közölnünk kell analitikus döntéseinket és a mellette szóló érveket
  - Ez az alapfeltétele az analitikus megismételhetőségnek: más kutató ugyanazokat az adatokat vizsgálva megegyező analitikus úton ugyanarra az eredményre érkezik

\* Ez sajnos ma a tudományban nem magától értetődő!

- Több valid elemzési út is lehet ugyanannak a kérdésnek a vizsgálatára
  - A szakmai közösség feladata ezeknek a felülbírálata
  - Lásd: multi analyst kutatások
- A tudomány iteratív módon lépésenként épül fel
  - Lehetséges, hogy a ma még megbízhatónak számító elemzési utat új eredmények fényében elvetjük
  - És ezzel nincs baj!

# 3 Adat

## 3.1 Mi az adat?

- Az **adat változók összessége**, amelyek valamilyen **mérés** eredményét rögzítik
  - Az **adat**, angolul **data** többesszám, egyesszámban **datum**
    - \* **Tehát data are és nem data is!**
- Sokszor gondolunk az adatra, mint egy adott, önmagában létező dologra. Ezzel szemben, az adat egy dolog mérés által létrejött reprezentációja
  - Tehát a mérés módja befolyásolja az adatok jelentését és minőségét
  - Illetve a mérés befolyásolja az adatokból levont következtetéseket is
  - Ismerhetünk akárhány statisztikai trükköt az adatok tisztítására, transzformálására, összesítésére, ha azok zajos, torzított, hibás mérésből származnak az adatokból statisztika útján levont következtetések is torzítani fognak
  - Ezért fontos már a kutatás megtervezésekor nagyon odafigyelni **mit** mérünk és **milyen pontossággal** tudjuk mérni.
  - Az adatok létrejöttével a kutatómódszertan foglalkozik
  - Példa:
    - \* 1999-ben a Mars Climate Orbiter a Mars atmoszférájába lépve darabokra hullott, mert az egyik modulja angolszász mértékegységeket használt, míg a másik metrikus
  - További olvasmányok:
    - \* <https://www.psychologicalscience.org/observer/measurement-matters>
    - \* <https://journals.sagepub.com/doi/full/10.1177/2515245920952393>
- Minden változónak legalább két különböző értéke kell legyen, hogy változónak nevezzük, különben egy konstans lenne
- A változóknak különböző típusai lehetnek, annak fényében milyen mérés eredményeit rögzítik