

Mit mond a XXI. század emberének a statisztika?

Rudas Tamás

Magyar Tudományos Akadémia
Társadalomtudományi Kutatóközpont
Eötvös Loránd Tudományegyetem
Statisztika Tanszék

Nehéz a jövőbe látni

Változik a statisztika gyakorlata és elmélete

Változnak a laikus és döntéshozói igények

Változik az adatokban való hit szintje

Változik az adat fogalma

Statisztika az ezredfordulóig

	Megfigyelés	Kísérlet
Populáció	összegzés (hivatalos statisztika, közgazdaságtudomány)	„természetes kísérlet” (politikatudomány)
Minta	felmérés, statisztikai következtetés (társadalomtudomány, piackutatás)	tervezett kísérlet (orvostudomány, biológia, pszichológia)

Minőségi követelmények

	Megfigyelés	Kísérlet
Populáció	pontosság, időbeliség, kis megterhelés	exogén vagy endogén kezelés
Minta	mintavétel, bekerülési valószínűségek, operacionalizálás	véletlenítés, szabálykövetés, kiesés, külső érvényesség

Total survey error

Teljes hiba = mintavételi hiba +
nem mintavételi hiba

survey value – census value – true value

érvényesség és megbízhatóság

Statisztika az ezredforduló után

Gyűjtött adatok	Megfigyelés	Kísérlet
Populáció	összegzés (hivatalos statisztika, közgazdaságtudomány)	„természetes kísérlet” (politikatudomány)
Minta	felmérés, statisztikai következtetés (társadalomtudomány, piackutatás)	tervezett kísérlet (orvostudomány, biológia, pszichológia)
Talált adatok	Megfigyelés	Kísérlet
Populáció	„bányászat”, „tanulás”	„természetes kísérlet”
Minta	önszelekció	önszelekció

Talált adatok

Big data (John Tukey, 1950?)

Process produced data

Organic data

Talált adatok

Adatgyűjtés és elemzés

Web scraping

Web harvesting

Web mining

Web farming

Data mining

Machine learning

Minőség talált adatoknál:

Sok korábbi minőségi kritérium érvényét veszti
(Conference of European Statistics Stakeholders, 2016)

De:

Érvényesség

Minta vagy populáció? (önszelekció)

Gyorsaság

Átláthatóság (vizualizáció)

Adatgyűjtés és adatelemzés

Hagyományos:

Elemzési terv

Adatgyűjtés

Adatelemzés

Elkülönült fázisok

Új:

Kérdés – adatgyűjtés – válasz

Egységes kezelés

Adatgenerálás helyett adatválogatás

Hagyományos:

Tervezett és költséges adatgyűjtés

Összeírás és felmérés

Új:

Állandó adatfolyam

Megbízható adatok kiválogatása

Ismételhető

Minőség és hit

Hagyományos:

Kidolgozott minőség fogalom

Eljárások a minőség javítására

Új:

Hit abban, hogy a Big Data minden
információt tartalmaz

Rendszeres adatszolgáltatás – folytonos adatszolgáltatás

Hagyományos:

Rendszeres, előre tervezett adatfelvételek

Új:

Bármikor végrehajtható lekérdezés

Például: tíz évenkénti census helyett az aktuális regisztertartalmak lekérdezése

Statisztika és szakpolitikai döntések

Hagyományos:

Megrendelés

Adat

Döntés

Adat

Új:

Interaktív

Real-time

Mesterséges intelligencia

Hagyományos:

Szakemberek által eldöntött kérdések

Adatszolgáltatás a felhasználóknak

Új:

Laikus kérdések

AI alapú adatkeresés

AI alapú válaszok

A számok és ábrák statisztikája helyett multimédia

Hagyományos:

Táblázatok

Összefoglaló mutatók

Ábrák

Új:

Multimédia interface

Statisztikai következtetés

Minta – bekerülési valószínűségek

Becslés, torzítatlanság, hatékonyság

Hipotézisvizsgálat, 1. és 2. fajú hiba

Valószínűségi jellegű mérési hibák kezelhetőek

Nem kezelhető: válaszhiány, hamis válaszok

Következtetés talált adatokból

Nincs elmélet

Minta vagy populáció

Exploratív elemzés

Elhisszük-e amit találunk

Alternatív következtetések

Kizárólagos igazság vagy konstruált igazságok

Okság

Jó, ha a szakpolitikai döntések felismert oksági kapcsolatokon alapulnak

Kísérlet – oksági kapcsolat

Megfigyeléses vizsgálat – új, részben elfogadott elméletek (Rubin, Pearl)

Okság és talált adatok

Nincs oksági elmélet talált adatokra

Talán Rubin elmélete alkalmazható

Causal effect, evidential effect

Hatás, amelynek létére bizonyíték van

Talált adatokból evidential effect

Elegendő?

Hasznos(?) statisztika

Tőzsdei adatok

Statisztikai elemzés

Befektetési tanácsok

Nyerő stratégiák

Óriási üzlet

De vajon az árfolyamok múltbeli mozgásából meghatározható a jövőbeli?

Szórakoztató(?) statisztika

Sport

Statisztikai elemzés

Nyerő mozdulatok

Nyerő stratégiák

Nyerő játékosok

De vajon mérhető és elemezhető mozdulatokból
összerakható-e egy nyerő stratégia

Houston Astros

ROVATOK LEGFRISSEBB IZ Totalcar JáAutórk Divány Femina Inda Otthontérkép Jobinfo Állás.hu

index HOLLYWOODI ZAKLATÁSOK SZÍNHÁZI ZAKLATÁSOK MAGYARORSZÁGON 2017. 11. 16. csütörtök Órán EUR: 311,89 Ft ▲ 1 °C GBP: 349,56 Ft ▼ 8 °C

BELFÖLD KÜLFÖLD GAZDASÁG TECH TUDOMÁNY KULT SPORT VÉLEMÉNY VIDEÓ FOTÓ 24 ÓRA

SPORT HOUSTON ASTROS MONKEYBALL BASKETBALL WORLD SERIES

A vegyész és a mérnök-pszichológus baseballtörténelmet írt



KÓRKONAI GERGELY
KÖVETÉS

2017.11.04. 00:02

Facebook share icon

2016-ban száznyelc éves sikeretlenséget és egy átkot megtörve lett baseballbajnok a Chicago Cubs. Kell ennél jobb sztori? Itt a 2017-es bajnok Houston Astrosé, ami 2013-ban még a bajnokság legrosszabb csapata volt, majd egy vegyész és egy légitechnikai gépészmérnök forradalmi képleteinek köszönhetően, végtelen számítógépes modellezés után bajnokok lettek. Ja, és annyira jól teszik a dolgukat, hogy van, akit 46 hónap börtönre ítélték, annyira irigy volt rájuk.

KAPCSOLÓDÓ CIKKEK (2)

- Először lett baseballbajnok a Houston Astros
- Csak megérette a kecskeáldozat és a felrobbantott labda

Összegzés

Paradigmaváltás zajlik