



1. rész

Statisztika és informatika

Mi a statisztika és miért fontos egy informatikusnak?

Komputerstatisztika kurzus

Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

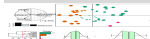
Irodalomjegyzék

Összefoglalás

Barczy Mátyás és Ispány Márton 2010
Informatikai Kar
Debreceni Egyetem

Az 1. rész témái

- 1 Mi a statisztika?
- 2 A statisztika eredete
- 3 A statisztikai munka felépítése
- 4 A statisztika módszerei
- 5 Számítógépes statisztika
- 6 Statisztika és információtechnológia



Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás



Mi a statisztika?

A statisztika
eredeteA statisztikai
munka felépítéseA statisztika
módszereiSzámítógépes
statisztikaStatisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás

Hol találkozunk a statisztikával a mindennapokban?

Meglepően gyakran:

- a médiában (tévéműsorok, újságok)
- az utcán (kérdezőbiztosok)
- az interneten (web–statisztikák)
- számos tudományterületen (pl. közgazdaságtan)

„Annak a statisztikának hiszek, melyet én hamisítottam”
(Winston Churchillnek tulajdonított mondás).

Meglepő lehet ez alapján, de a statisztika egy **tudomány!**

A statisztika heurisztikus definíciója

Megfigyelésekből, kísérletekből származó **adatok** elemzése, azokból következtetések levonása.



Mit értünk statisztika alatt?

A statisztika szó többféle jelentéssel bír:

- A statisztika mint tudomány.
- A statisztika mint gyakorlati tevékenység.
- A statisztika mint mutatószám.
- A statisztika mint leképezés.

Példa (Támogatja-e a halálbüntetés visszaállítását?)

A lakosság véleményére vagyunk kíváncsiak a halálbüntetés támogatása illetve ellenzése kérdésében.

A megbízható eredményhez a statisztika tudománya által megalapozott módszereket kell alkalmazni.

A felmérést közvéleménykutató céggel végeztethetjük el. Az eredmény egy statisztikai mutatószám: a lakosság hány százaléka ellenzi a halálbüntetést.

[Mi a statisztika?](#)

[A statisztika eredete](#)

[A statisztikai munka felépítése](#)

[A statisztika módszerei](#)

[Számítógépes statisztika](#)

[Statisztika és információtechnológia](#)

[Irodalomjegyzék](#)

[Összefoglalás](#)



A statisztika jelentés szerinti definíciói

A statisztika mint tudomány

A **valóság** tényeinek sokaságát, tömegjelenségeket számszerű, tömör formában leíró tudomány.

A statisztika mint tevékenység

Az a **gyakorlati** tevékenység, az adatgyűjtéstől a végső mutatószámok előállításáig, amely során a fenti leírást elvégezzük.

A statisztika mint fogalom

Minden összesítés, származtatott szám, illetve az alapadatokból ezt a számot előállító leképezés.

Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás

Történeti háttér: kezdetek

- Sumérok, i.e. 3000: a lakosság összeírása ékírásos táblákon.
- Egyiptom, II. Amasis (Ahmose II) fáraó, 26. dinasztia, i.e. 570–526: Egyiptom utolsó nagy fáraója, aki intézményesítette a népesség összeírását (a halálozás követése).
- Kína, Yao császár, i.e. 2358–2258: legendás kínai császár, a népsűrűség számolása.
- India, Arthashastra, i.e. 4. század: értekezés az államigazgatásról, a gazdaságról és a hadtudományokról. (Csandragupta király (i.e. 313–289) Kautilya nevű miniszterét tartják fő szerzőjének.)



Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás

II. Amasis fáraó és Yao császár



Copyright: Egyptian Museum of Berlin, szobortöredék



Copyright: Wikipedia, selyem festmény



Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás



Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás

A statisztika szó eredete

A középkorban statisztika alatt államtudományt, az állammal kapcsolatos ismeretek összességét értették:

status = állam (latinul)

Girolamo Ghilini

Olasz tudós, aki **Teatro d'Huomi Letterati** c. munkáját 1589–ben az alábbi szavakkal jellemezte:

civile, politica, statistica e militare scienza

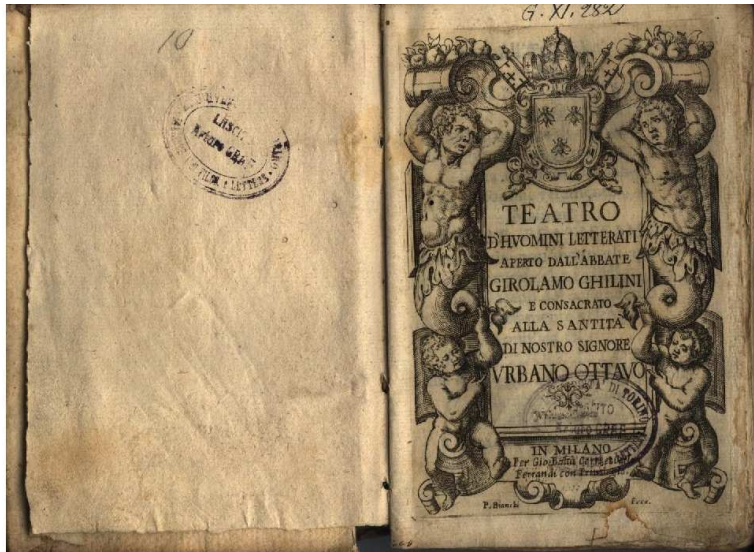
A műben kb. 5000 itáliai tudós, művész, politikus életrajza szerepel.

Gottfried Achenwall (1719–1772)

Német udvari kancellár, jogtudós, filozófus és statisztikus. Az alábbi művében szintén előfordul a statisztika szó:

Staatsverfassung der heutigen vornehmsten europäischen Reiche und Völker im Grundrisse

Girolamo Ghilini munkája



Copyright: Biblioteca Centrale della Facolt'a di Lettere e Filosofia, Università degli Studi di Torino

1. rész

© Barczy Máttyás
és Ispány Márton
2010

Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás

[Mi a statisztika?](#)[A statisztika
eredete](#)[A statisztikai
munka felépítése](#)[A statisztika
módszerei](#)[Számítógépes
statisztika](#)[Statisztika és infor-
mációtechnológia](#)[Irodalomjegyzék](#)[Összefoglalás](#)

Az angol politikai számtan

A mai értelemben használt statisztika kezdete.

John Graunt (1620–1674)

Az első demográfus, epidemológus (eredeti foglalkozása: rőfös, mai szóhasználattal méteráru-kereskedő), a Royal Society tagja. Fő műve:

*Natural and Political **Observations** Made
upon the Bills of Mortality (1662)*

London lakosságának első felmérése, statisztika az előforduló betegségekről.

[Mi a statisztika?](#)[A statisztika
eredete](#)[A statisztikai
munka felépítése](#)[A statisztika
módszerei](#)[Számítógépes
statisztika](#)[Statisztika és infor-
mációtechnológia](#)[Irodalomjegyzék](#)[Összefoglalás](#)

Sir William Petty (1623–1687)

Közgazdász és filozófus, a Royal Society tagja. Cromwell alatt az elfoglalt Írországot mérte fel, számos közgazdasági fogalmat vezetett be, a népszámlálás (cenzus) fontosságának hangsúlyozója. Egyik fő műve:

A Treatise of Taxes and Contributions (1662)

Szisztematikusan használta az **átlagolás műveletét**, bevezette a becslés fogalmát. Megbecsülte London lakosságát az export és a halálozás alapján. 30%-os növekedés az exportban ugyanannyi növekedést jelent a lakosságban (**regresszió**). A halálozások számát 30-cal szorozva határozta meg a lakosság számát.

Graunt és Petty



CAPTAIN JOHN GRAUNT

Copyright: ismeretlen művész
rajza



Copyright: Wikipedia

Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás

[Mi a statisztika?](#)[A statisztika
eredete](#)[A statisztikai
munka felépítése](#)[A statisztika
módszerei](#)[Számítógépes
statisztika](#)[Statisztika és infor-
mációtechnológia](#)[Irodalomjegyzék](#)[Összefoglalás](#)

Elméleti alapok: valószínűségszámítás

- Kocka– és kártyajátékok: szerencsejáték, jóslás. A lehetséges kimenetek összeszámlálása, a relatív gyakoriság fogalmának kialakulása.
- Blaise Pascal (1623–1662) és Daniel Bernoulli (1700–1782) a valószínűségszámítás kezdetleges megalapozása a szerencsejátékok vizsgálatával.
- Adolphe Quetelet (1796-1874): átlagos ember és normális görbe. Első statisztikai kongresszus 1853–ban.
- Pierre-Simon Laplace márki (1749–1827), Carl Friedrich Gauss (1777–1855) és Adrien-Marie Legendre (1752–1833): legkisebb négyzetek módszere (méthode des moindres carré, 1806) és a Gauss–Laplace (normális) eloszlás.
- A. N. Kolmogorov (1903–1987): a valószínűségszámítás elméleti (axiomatikus) megalapozása.



Tudományos megismerés

A valóság megfigyeléséből absztrakció útján törvényeket, modelleket alkotunk, majd ezeket szembesítjük a valósággal ellenőrizve őket.

- 1 Megfigyelés
- 2 Elemzés
- 3 Modellalkotás
- 4 Előrejelzés

A fenti lépéseket gyakran **ciklikusan** alkalmazzuk.

Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás

1. Megfigyelés

Ebben a fázisban hozzuk létre a statisztikai adatmezőt, gyakran egyszerűen mintavételről, adatgyűjtésről beszélünk.

Nagy mértékben kihat arra, hogy később milyen elemzéseket végezhetünk. A rosszul kialakított adatmező megköti az elemző kezét, lehet, hogy éppen a vizsgálat által megcélzott kérdésekre nem tudunk majd válaszolni.

Adatminőség, adattisztítás.

Hasznos, ha az elemző már az adatmező kialakításában is részt vesz, már ekkor átgondolva azt, hogy a feldolgozás során milyen módszereket lehet és érdemes alkalmazni.

Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

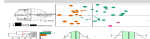
Összefoglalás

2. Elemzés

Szabályosságok, összefüggések, mintázatok keresése az adatállományban.

Ebben a fázisban szokás **módosítani**, esetleg **átkódolni** az adatokat. Itt merülhet fel a **hiányzó** és a **kilógó** adatok problémája (**missing values** és **outliers**).

Az itt alkalmazott módszerek általában az **exploratív** (feltáró jellegű) adatelemzés körébe tartoznak. Például, leíró statisztikák, grafikus megjelenítések, többdimenziós skálázás.



[Mi a statisztika?](#)

[A statisztika eredete](#)

[A statisztikai munka felépítése](#)

[A statisztika módszerei](#)

[Számítógépes statisztika](#)

[Statisztika és információtechnológia](#)

[Irodalomjegyzék](#)

[Összefoglalás](#)

3. Modellalkotás

Az elemzési fázisban talált szabályosságok, összefüggések, mintázatok matematikai statisztikai modellekkel való leírása.

Először magát a modellt alakítjuk ki, majd ellenőrizzük alapfeltevéseit és megbecsüljük paramétereit.

Itt ellenőrizzük azt is, hogy a választott modell mennyire jól illeszkedik az adatokra, esetleg több illesztett modell közül választjuk ki a legjobbat.

Ezen fázisbeli módszerek legtöbbször a **konfirmatív** (igazolójellegű) adatelemzés körébe tartozik.



Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás

4. Előrejelzés

A megalkotott modell alapján **előrejelzés**, **becslés** készítése. Ezt sokszor sztochasztikus szimulációval hajtjuk végre. (**Monte Carlo módszerek**)
A szimulációs eredményeket új megfigyelésekkel szembevetve a modell érvényességének korlátai vizsgálhatóak.



Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás

A népesség súlyának (IQ-jának stb.) becslése:

- Egy kis, reprezentatív részt kiválasztva mintát veszünk.
- A minta vizsgálatával (pl. grafikus módszerekkel) próbáljuk megsejteni, hogy a vizsgált jellemző milyen eloszlást követhet.
- Tegyük fel, hogy a vizsgálatok (pl. a minta hisztogramja) normális eloszlásra utalnak. Ezen eloszlásnak két ún. paramétere van: az átlag (várható érték) és a szórás. Ezzel ún. paraméteres feladattá tudjuk redukálni a problémát.
- Valamilyen módszerrel (pl. maximum likelihood) elvégezzük a paraméterek becslését a minta alapján.
- Előrejelzésként ún. konfidencia intervallumot szerkesztünk, amely a népesség 95%-át tartalmazza. Ezt aztán új adatokon ellenőrizzük. Nagy hiba esetén kezdjük újra az eljárást.



[Mi a statisztika?](#)

[A statisztika eredete](#)

[A statisztikai munka felépítése](#)

[A statisztika módszerei](#)

[Számítógépes statisztika](#)

[Statisztika és információtechnológia](#)

[Irodalomjegyzék](#)

[Összefoglalás](#)



[Mi a statisztika?](#)

[A statisztika
eredete](#)

[A statisztikai
munka felépítése](#)

[A statisztika
módszerei](#)

[Számítógépes
statisztika](#)

[Statisztika és infor-
mációtechnológia](#)

[Irodalomjegyzék](#)

[Összefoglalás](#)

A statisztika hidat alkot az elméleti matematika és más (természet-, illetve társadalom-) tudományok között. Ezek a tudományok a legtöbb matematikai módszert ugyanis a statisztikán keresztül alkalmazzák. A gyakorlati, alkalmazott statisztika elméleti háttérét a matematikai statisztika adja, amely felépítésében a szokásos matematikai gondolkodást követi, azaz definíció, tétel és bizonyítás, és a kalkulus illetve főként a valószínűségszámítás módszereire támaszkodik.



Alkalmazott (számítógépes) statisztika

Csak az a statisztika hasznos, melynek valós alkalmazása is van.

Egy statisztikai eljárást kétféle úton hajthatunk végre számítógép segítségével

- Egy általunk ismert programozási nyelven (Pascal, C, Java, Haskell stb.) leprogramozzuk.
- Statisztikai (matematikai) programcsomagot használunk (lásd később).

Mi a statisztika?

A statisztika eredete

A statisztikai munka felépítése

A statisztika módszerei

Számítógépes statisztika

Statisztika és információtechnológia

Irodalomjegyzék

Összefoglalás



Mit kell tartalmaznia egy statisztikára alkalmas programcsomagnak?

- mátrixnyelv: műveletek mátrixokkal, invertálás, lineáris egyenletrendszer megoldása.
- függvénykönyvtár: statisztikai és matematikai függvények.
- grafikus felület: alapvető statisztikai ábrák, diagramok.

Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás



- Szabad szoftver: az R nyelv és rendszer.
- Statisztikai szoftverek: SPSS, SAS, STATA, S-Plus, Statistica, SigmaStat.
- Általános célú matematikai szoftverek: MATLAB (SciLab), Mathematica, Maple.

Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás



A modern információtechnológia (IT) számos statisztikai módszert, eljárást használ, illetve az IT szoftverek mögötti elméleti háttérrel sokszor a statisztika nyújtja.

Információtechnológia

Az információk (adatok) számítástechnikai eszközökkel történő összegyűjtésének, tárolásának, feldolgozásának és megjelenítésének a módszertana.

Az információ feldolgozás piramisa:

- Adatbázisok
- Vállalatirányítási rendszerek
- Üzleti intelligencia rendszerek

[Mi a statisztika?](#)

[A statisztika eredete](#)

[A statisztikai munka felépítése](#)

[A statisztika módszerei](#)

[Számítógépes statisztika](#)

[Statisztika és információtechnológia](#)

[Irodalomjegyzék](#)

[Összefoglalás](#)



- Adatintegráció
- ETL folyamat (Extracting, Transforming, Loading): az adatok különböző rendszerekből történő elérésének, transzformációjának, integrációjának és betöltésének automatizált működtetése.
- CRM (Customer Relationship Management): ügyfélkapcsolat-menedzsment
- IT-menedzsment, hatékonyság-menedzsment
- Rizikó–kezelés
- Vezetői információs rendszerek

Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás



Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás

① M. G. Kendall, *Biometrika* (47) 447–449, 1960.



Mi a statisztika?

A statisztika
eredete

A statisztikai
munka felépítése

A statisztika
módszerei

Számítógépes
statisztika

Statisztika és infor-
mációtechnológia

Irodalomjegyzék

Összefoglalás

- 1 A statisztika az adatelemzés, feldolgozás módszertana.
- 2 A statisztika történeti háttere.
- 3 Statisztikai eljárások megvalósítása számítógépen.
- 4 A statisztika fontos szerepet játszik a haladott információtechnológiai eljárásoknál.