

# Adatfeldolgozási alapismeretek

Tóthmérész Béla

Blackbox for Windows

# Néhány alapigazság (1)

- Bárki lehet kiváló statisztikus különösebb matematikai előképzettség nélkül
- A tudás mértéke a befektetett idővel és energiával arányos (az összefüggés sajnos nem lineáris!)

## Néhány alapigazság (2)

- Ekkora óraszámban nem lehet a tárgynak még az alapjait sem elsajátítani, legfeljebb „megismerkedni” vele
- Óraszám és erőfeszítés hiányában kénytelenek leszünk a fejünket használni – ***ezt többféleképpen is megtehetjük !***

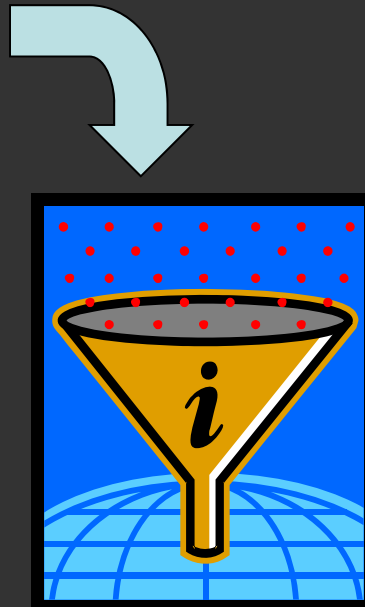


Blackbox for Windows

Használd a fejed !

# Miért van szükség statisztikára?

445 446 397 226  
388 3445 188 1002  
47762 432 54 12  
98 345 2245 8839  
77492 472 565 999  
1 34 882 545 4022  
827 572 597 364



Az adatokból  
hasznosítható  
információt nyerjünk



$$\bar{x} = \dots$$

# Miért van szükség statisztikára?

Zaj, zaj, zaj, zaj, zaj !



OK - lehet, hogy ennyire azért nem.

# Mi a statisztika?

“A lehetetlen dolgok rendszerint nem történnek meg.”

- *Sam Treiman, Princeton University*

- *A statisztika segít számszerűsíteni a “rendszerint”-et.*

# Mi a statisztika?

Általános értelemben:

Az adatok gyűjtésének és feldolgozásának tudománya.

Szűkebb értelemben:

A matematika egy részterülete . . .

???



# Mi a statisztika?

- “Lies, damn lies, and statistics!”  
*Moroney*
- “a valóság mennyiségi összefüggéseit, számadatokkal kifejezhető tényeit vizsgáló tudomány, ill. gyakorlati tevékenység.”  
*Britannica Hungarica*

# Mi a statisztika?

“Statisztikának nevezik az ilyen tevékenység eredményeként keletkező adathalmazt, ill. adatrendszeret is.”

*Britannica Hungarica*

# Az előadások célja

- Intuitív áttekintést adjon a modern statisztikáról, elsősorban a robusztus, adatközpontú részéről.
- Nem célja bonyolult képletek céltalan megtanítása
- Egyfajta szellemi kaland
- Szép emlék, mint az első szerelem

# Stochasztikus diszciplínák



# Statisztika



**Klasszikus  
(Fisherianus)**

Ronald Aylmer Fisher

**Nem klasszikus  
(Tukey-ánus)**

John Wilder Tukey

# Klasszikus statisztika

Számos *feltételt támaszt* az adatokkal szemben, amelyek igen gyakran csak *feltételezhetően teljesülnek*.

Kevésbé „szemérmesen” fogalmazva ez azt jelenti, hogy *NEM*!

# Nem klasszikus statisztika

Az adatokat *önmagában is értékesnek* tekinti és mindenféle körítés nélkül is értékelhetőnek tartja őket.

*EDA: Exploratory Data Analysis*

# Sir Ronald Aylmer Fisher

(1890-1962)

- Kiváló (evolúció) biológus
- Modern matematikai statisztika atyja
- Kisminintás módszerek
- Számos minta-statisztika eloszlásának pontos kidolgozója
- Varianciaanalízis megalkotója



Blackbox for windows



# John Wilder Tukey

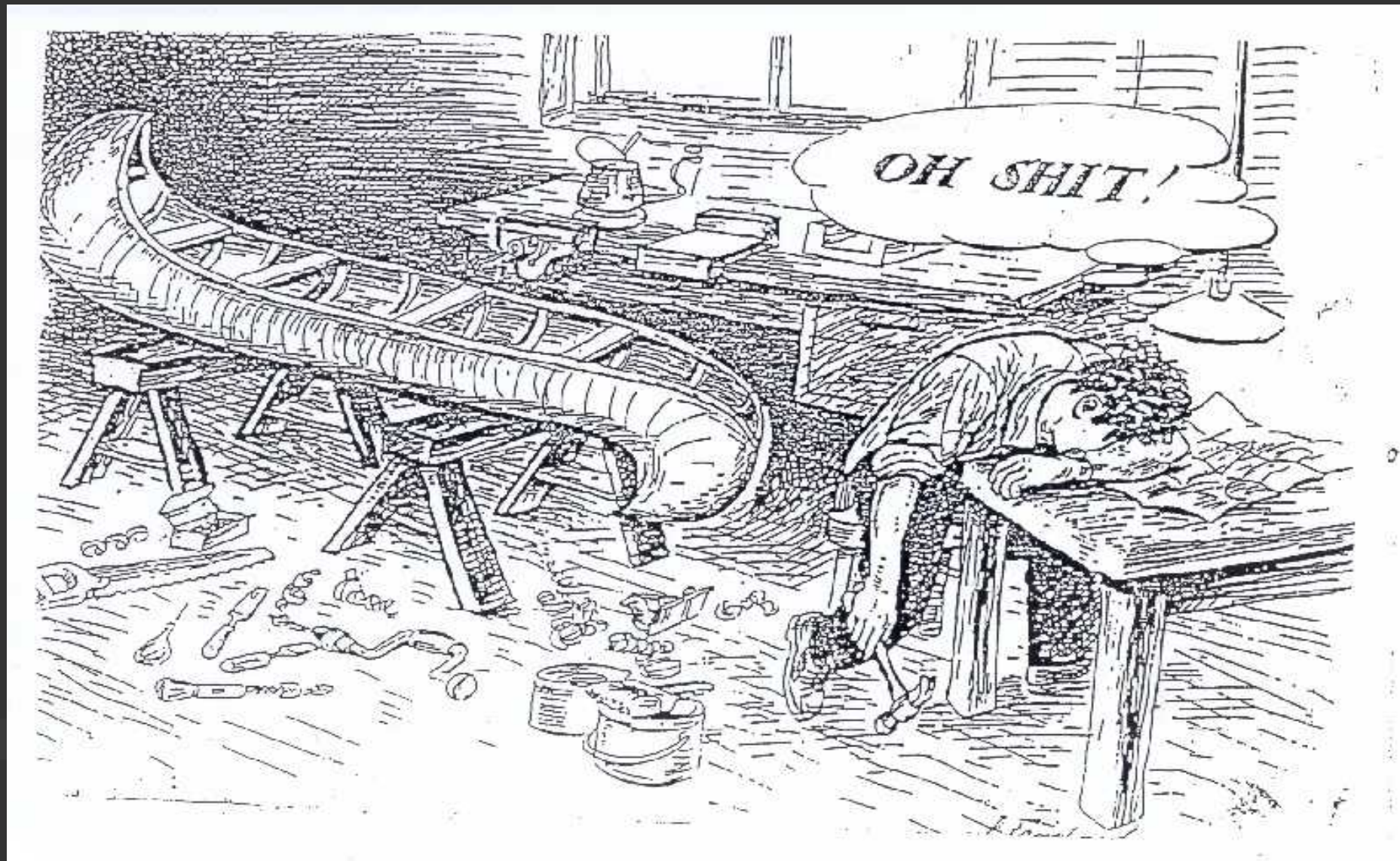
(1915-2000)

- EDA: tényfeltáró adatelemzés
- Robusztus módszerek
- Adatelemzés hangsúlyozása
- Grafikus módszerek reneszánsza
- Ő maga is egy reneszánsz ember



Blackbox for Windows

Ha megvan a megfelelő eszköz az még nem garancia ...!



# Jelölések, a formális apparátus bevezetése

# Helyzetparaméter becslése

# Szóródási (skála) paraméter becslése (variabilitás)

# R programnyelv és programozási környezet

# Grafikus módszerek

Blackbox for windows