

# p-érték, hipotézistesztesztelés, és ellentmondásaik

Ferenci Tamás

tamas.ferenci@medstat.hu

<http://www.medstat.hu/>

<https://www.youtube.com/c/FerenciTamas>

Utoljára frissítve: 2022. június 30.

# Következtésetelmélet

- A megfigyelt világ és a tudásunk összekapcsolása
- Deduktív következtetés: kiindulunk egy hipotézisből (amit a tudásunk alapján fogalmaztunk meg), és ebből következtetünk arra, hogy mit fogunk látni a világban
  - Objektív (abban az értelemben, hogy mindig igaz, ha a kiinduló hipotézis igaz volt)
  - Nem alkalmazható a tudásunk bővítésére
- Induktív következtetés: az alapján, hogy mit látunk a világban, következtetünk arra, hogy mely hipotézis lehet igaz
  - A tudásunk bővítése
  - Az objektív tudásunk szűkebb köré
  - Amire az objektív világ látszik lenni
- Számos kísérlet ennek feloldására (pl. Popper és a falszifikáció → megspórolható az indukció; ezt valósítja meg a hipotézisvizsgálat is, mindkét iskola)

Goodman SN. Toward evidence-based medical statistics. 1: The P-value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# Következtésetelmélet

- A megfigyelt világ és a tudásunk összekapcsolása
- Deduktív következtetés: kiindulunk egy hipotézisből (amit a tudásunk alapján fogalmaztunk meg), és ebből következtetünk arra, hogy mit fogunk látni a világban
  - Objektív (abban az értelemben, hogy mindig igaz, ha a kiinduló hipotézis igaz volt)
  - Nem alkalmazható tudásunk bővítésére
- Induktív következtetés: az alapján, hogy mit látunk a világban, következtetünk arra, hogy mely hipotézis lehet igaz
  - Nem alkalmazható tudásunk bővítésére
  - A tudásunk bővítésére azért nem alkalmas, mert a tudásunk alapján fogalmaztunk meg hipotézist, és ebből következtetünk arra, hogy mit fogunk látni a világban
  - A tudásunk bővítésére azért nem alkalmas, mert a tudásunk alapján fogalmaztunk meg hipotézist, és ebből következtetünk arra, hogy mit fogunk látni a világban
- Számos kísérlet ennek feloldására (pl. Popper és a falszifikáció → megspórolható az indukció; ezt valósítja meg a hipotézisvizsgálat is, mindkét iskola)

Goodman SN. Toward evidence-based medical statistics. 1: The P-value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# Következtésetelmélet

- A megfigyelt világ és a tudásunk összekapcsolása
- Deduktív következtetés: kiindulunk egy hipotézisből (amit a tudásunk alapján fogalmaztunk meg), és ebből következtetünk arra, hogy mit fogunk látni a világban
  - Objektív (abban az értelemben, hogy mindig igaz, ha a kiinduló hipotézis igaz volt)
    - Nem alkalmazható tudásunk bővítésére
- Induktív következtetés: az alapján, hogy mit látunk a világban, következtetünk arra, hogy mely hipotézis lehet igaz
  - Nem objektív, mert a tudásunk függ a megfigyelésünktől
  - Nem bizonyítható, mert a tudásunk soha nem lesz teljes
  - Az objektív következtetés nem alkalmazható
- Számos kísérlet ennek feloldására (pl. Popper és a falszifikáció → megspórolható az indukció; ezt valósítja meg a hipotézisvizsgálat is, mindkét iskola)

Goodman SN. Toward evidence-based medical statistics. 1: The P-value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# Következtésteória

- A megfigyelt világ és a tudásunk összekapcsolása
- Deduktív következtetés: kiindulunk egy hipotézisből (amit a tudásunk alapján fogalmaztunk meg), és ebből következtetünk arra, hogy mit fogunk látni a világban
  - Objektív (abban az értelemben, hogy mindig igaz, ha a kiinduló hipotézis igaz volt)
  - Nem alkalmazható tudásunk bővítésére
- Induktív következtetés: az alapján, hogy mit látunk a világban, következtetünk arra, hogy mely hipotézis lehet igaz
  - A bizonyíték fogalma
  - A tudásunk és a bizonyítékunk közötti kapcsolat
  - A bizonyíték és a hipotézis közötti kapcsolat
- Számos kísérlet ennek feloldására (pl. Popper és a falszifikáció → megspórolható az indukció; ezt valósítja meg a hipotézisvizsgálat is, mindkét iskola)

Goodman SN. Toward evidence-based medical statistics. 1: The P-value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# Következtéselmélet

- A megfigyelt világ és a tudásunk összekapcsolása
- Deduktív következtetés: kiindulunk egy hipotézisből (amit a tudásunk alapján fogalmaztunk meg), és ebből következtetünk arra, hogy mit fogunk látni a világban
  - Objektív (abban az értelemben, hogy mindig igaz, ha a kiinduló hipotézis igaz volt)
  - Nem alkalmazható tudásunk bővítésére
- Induktív következtetés: az alapján, hogy mit látunk a világban, következtetünk arra, hogy mely hipotézis lehet igaz
  - A bizonyíték fogalma
  - Új tudásra tudunk szert tenni
  - Ám ez soha nem tud biztos lenni
- Számos kísérlet ennek feloldására (pl. Popper és a falszifikáció → megspórolható az indukció; ezt valósítja meg a hipotézisvizsgálat is, mindkét iskola)

Goodman SN. Toward evidence-based medical statistics. 1: The P-value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# Következtésetelmélet

- A megfigyelt világ és a tudásunk összekapcsolása
- Deduktív következtetés: kiindulunk egy hipotézisből (amit a tudásunk alapján fogalmaztunk meg), és ebből következtetünk arra, hogy mit fogunk látni a világban
  - Objektív (abban az értelemben, hogy mindig igaz, ha a kiinduló hipotézis igaz volt)
  - Nem alkalmazható tudásunk bővítésére
- Induktív következtetés: az alapján, hogy mit látunk a világban, következtetünk arra, hogy mely hipotézis lehet igaz
  - A bizonyíték fogalma
  - Új tudásra tudunk szert tenni
  - Ám ez soha nem tud biztos lenni
- Számos kísérlet ennek feloldására (pl. Popper és a falszifikáció → megspórolható az indukció; ezt valósítja meg a hipotézisvizsgálat is, mindkét iskola)

Goodman SN. Toward evidence-based medical statistics. 1: The P-value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# Következtésteória

- A megfigyelt világ és a tudásunk összekapcsolása
- Deduktív következtetés: kiindulunk egy hipotézisből (amit a tudásunk alapján fogalmaztunk meg), és ebből következtetünk arra, hogy mit fogunk látni a világban
  - Objektív (abban az értelemben, hogy mindig igaz, ha a kiinduló hipotézis igaz volt)
  - Nem alkalmazható tudásunk bővítésére
- Induktív következtetés: az alapján, hogy mit látunk a világban, következtetünk arra, hogy mely hipotézis lehet igaz
  - A bizonyíték fogalma
  - Új tudásra tudunk szert tenni
  - Ám ez soha nem tud biztos lenni
- Számos kísérlet ennek feloldására (pl. Popper és a falszifikáció → megspórolható az indukció; ezt valósítja meg a hipotézisvizsgálat is, mindkét iskola)

Goodman SN. Toward evidence-based medical statistics. 1: The P-value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.



# Következtésteória

- A megfigyelt világ és a tudásunk összekapcsolása
- Deduktív következtetés: kiindulunk egy hipotézisből (amit a tudásunk alapján fogalmaztunk meg), és ebből következtetünk arra, hogy mit fogunk látni a világban
  - Objektív (abban az értelemben, hogy mindig igaz, ha a kiinduló hipotézis igaz volt)
  - Nem alkalmazható tudásunk bővítésére
- Induktív következtetés: az alapján, hogy mit látunk a világban, következtetünk arra, hogy mely hipotézis lehet igaz
  - A bizonyíték fogalma
  - Új tudásra tudunk szert tenni
  - Ám ez soha nem tud biztos lenni
- Számos kísérlet ennek feloldására (pl. Popper és a falszifikáció → megspórolható az indukció; ezt valósítja meg a hipotézisvizsgálat is, mindkét iskola)

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# Következtésteória

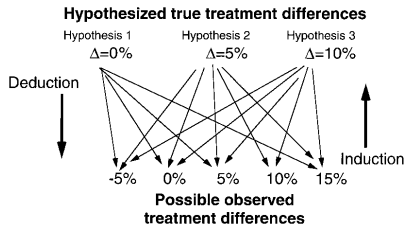
- A megfigyelt világ és a tudásunk összekapcsolása
- Deduktív következtetés: kiindulunk egy hipotézisből (amit a tudásunk alapján fogalmaztunk meg), és ebből következtetünk arra, hogy mit fogunk látni a világban
  - Objektív (abban az értelemben, hogy mindig igaz, ha a kiinduló hipotézis igaz volt)
  - Nem alkalmazható tudásunk bővítésére
- Induktív következtetés: az alapján, hogy mit látunk a világban, következtetünk arra, hogy mely hipotézis lehet igaz
  - A bizonyíték fogalma
  - Új tudásra tudunk szert tenni
  - Ám ez soha nem tud biztos lenni
- Számos kísérlet ennek feloldására (pl. Popper és a falszifikáció → megspórolható az indukció; ezt valósítja meg a hipotézisvizsgálat is, mindkét iskola)

Goodman SN. Toward evidence-based medical statistics. 1: The P-value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

- A megfigyelt világ és a tudásunk összekapcsolása
- Deduktív következtetés: kiindulunk egy hipotézisből (amit a tudásunk alapján fogalmaztunk meg), és ebből következtetünk arra, hogy mit fogunk látni a világban
  - Objektív (abban az értelemben, hogy mindig igaz, ha a kiinduló hipotézis igaz volt)
  - Nem alkalmazható tudásunk bővítésére
- Induktív következtetés: az alapján, hogy mit látunk a világban, következtetünk arra, hogy mely hipotézis lehet igaz
  - A bizonyíték fogalma
  - Új tudásra tudunk szert tenni
  - Ám ez soha nem tud biztos lenni
- Számos kísérlet ennek feloldására (pl. Popper és a falszifikáció → megspórolható az indukció; ezt valósítja meg a hipotézisvizsgálat is, mindkét iskola)

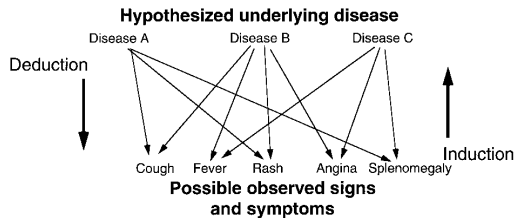
Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# Statisztikai következtetéselmélet



Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# „Orvosi következtetéselmélet”



Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# Két iskola ütközése: R. A. Fisher

- $p$ -érték: „measure of evidence”
- Mégpedig *egyetlen* hipotézisre vonatkozóan (tehát nincs ellenhipotézis)
- *Egyetlen* kísérletből akarunk dolgozni (nincs „képzeltbeli ismételt mintavétel”)
- *Nem* akarunk külső információt felhasználni (nincs prior valószínűség)
- *Nem* hibavalószínűség
- Hogyan is lehetne valószínűség → ahhoz több kísérletet kellene, legalább képzeletben, tekintenünk (ahogy annak a kijelentésnek, hogy „ez az érme 40% valószínűséggel ad fejet” is csak úgy van értelme, ha többször feldobhatjuk – ez a valószínűség frekvencionista értelmezése)
- Inkább informális index, esetleg az a szabály, de ez sem mechanikusan, hogy adott  $p$  alatt elutasítunk: „significance testing” (*nem* „hypothesis testing”!)

## Két iskola ütközése: R. A. Fisher

- $p$ -érték: „measure of evidence”
- Mégpedig *egyetlen* hipotézisre vonatkozóan (tehát nincs ellenhipotézis)
- *Egyetlen* kísérletből akarunk dolgozni (nincs „képzeletbeli ismételt mintavétel”)
- *Nem* akarunk külső információt felhasználni (nincs prior valószínűség)
- *Nem* hibavalószínűség
- Hogyan is lehetne valószínűség → ahhoz több kísérletet kellene, legalább képzeletben, tekintenünk (ahogy annak a kijelentésnek, hogy „ez az érme 40% valószínűséggel ad fejet” is csak úgy van értelme, ha többször feldobhatjuk – ez a valószínűség frekvencionista értelmezése)
- Inkább informális index, esetleg az a szabály, de ez sem mechanikusan, hogy adott  $p$  alatt elutasítunk: „significance testing” (*nem* „hypothesis testing”!)

## Két iskola ütközése: R. A. Fisher

- $p$ -érték: „measure of evidence”
- Mégpedig *egyetlen* hipotézisre vonatkozóan (tehát nincs ellenhipotézis)
- *Egyetlen* kísérletből akarunk dolgozni (nincs „képzeltbeli ismételt mintavétel”)
- *Nem* akarunk külső információt felhasználni (nincs prior valószínűség)
- *Nem* hibavalószínűség
- Hogyan is lehetne valószínűség → ahhoz több kísérletet kellene, legalább képzeletben, tekintenünk (ahogy annak a kijelentésnek, hogy „ez az érme 40% valószínűséggel ad fejet” is csak úgy van értelme, ha többször feldobhatjuk – ez a valószínűség frekvencionista értelmezése)
- Inkább informális index, esetleg az a szabály, de ez sem mechanikusan, hogy adott  $p$  alatt elutasítunk: „significance testing” (*nem* „hypothesis testing”!)



## Két iskola ütközése: R. A. Fisher

- $p$ -érték: „measure of evidence”
- Mégpedig *egyetlen* hipotézisre vonatkozóan (tehát nincs ellenhipotézis)
- *Egyetlen* kísérletből akarunk dolgozni (nincs „képzeltbeli ismételt mintavétel”)
- *Nem* akarunk külső információt felhasználni (nincs prior valószínűség)
- *Nem* hibavalószínűség
- Hogyan is lehetne valószínűség → ahhoz több kísérletet kellene, legalább képzeletben, tekintenünk (ahogy annak a kijelentésnek, hogy „ez az érme 40% valószínűséggel ad fejet” is csak úgy van értelme, ha többször feldobhatjuk – ez a valószínűség frekvencionista értelmezése)
- Inkább informális index, esetleg az a szabály, de ez sem mechanikusan, hogy adott  $p$  alatt elutasítunk: „significance testing” (*nem* „hypothesis testing”!)

## Két iskola ütközése: R. A. Fisher

- $p$ -érték: „measure of evidence”
- Mégpedig *egyetlen* hipotézisre vonatkozóan (tehát nincs ellenhipotézis)
- *Egyetlen* kísérletből akarunk dolgozni (nincs „képzeltbeli ismételt mintavétel”)
- *Nem* akarunk külső információt felhasználni (nincs prior valószínűség)
- *Nem* hibavalószínűség
- Hogyan is lehetne valószínűség → ahhoz több kísérletet kellene, legalább képzeletben, tekintenünk (ahogy annak a kijelentésnek, hogy „ez az érme 40% valószínűséggel ad fejet” is csak úgy van értelme, ha többször feldobhatjuk – ez a valószínűség frekvencionista értelmezése)
- Inkább informális index, esetleg az a szabály, de ez sem mechanikusan, hogy adott  $p$  alatt elutasítunk: „significance testing” (*nem* „hypothesis testing”!)

## Két iskola ütközése: R. A. Fisher

- $p$ -érték: „measure of evidence”
- Mégpedig *egyetlen* hipotézisre vonatkozóan (tehát nincs ellenhipotézis)
- *Egyetlen* kísérletből akarunk dolgozni (nincs „képzeltbeli ismételt mintavétel”)
- *Nem* akarunk külső információt felhasználni (nincs prior valószínűség)
- *Nem* hibavalószínűség
- Hogyan is lehetne valószínűség → ahhoz több kísérletet kellene, legalább képzeletben, tekintenünk (ahogy annak a kijelentésnek, hogy „ez az érme 40% valószínűséggel ad fejet” is csak úgy van értelme, ha többször feldobhatjuk – ez a valószínűség frekvencionista értelmezése)
- Inkább informális index, esetleg az a szabály, de ez sem mechanikusan, hogy adott  $p$  alatt elutasítunk: „significance testing” (*nem* „hypothesis testing”!)

## Két iskola ütközése: R. A. Fisher

- $p$ -érték: „measure of evidence”
- Mégpedig *egyetlen* hipotézisre vonatkozóan (tehát nincs ellenhipotézis)
- *Egyetlen* kísérletből akarunk dolgozni (nincs „képzeltbeli ismételt mintavétel”)
- *Nem* akarunk külső információt felhasználni (nincs prior valószínűség)
- *Nem* hibavalószínűség
- Hogyan is lehetne valószínűség → ahhoz több kísérletet kellene, legalább képzeletben, tekintenünk (ahogy annak a kijelentésnek, hogy „ez az érme 40% valószínűséggel ad fejet” is csak úgy van értelme, ha többször feldobhatjuk – ez a valószínűség frekvencionista értelmezése)
- Inkább informális index, esetleg az a szabály, de ez sem mechanikusan, hogy adott  $p$  alatt elutasítunk: „significance testing” (*nem* „hypothesis testing”!)

# Két iskola ütközése: J. Neyman és E. Pearson

- Döntés: „hibavalószínűség”
  - Két hipotézis között döntünk
  - Kísérletek sorát tételezzük fel...
  - ...és a célunk, hogy *hosszú távon* biztosítsunk adott első- és másodfajú hibaarányt
  - Egy adott, konkrét hipotézisről úgysem dönthető el, hogy mennyire igaz, a kérdés csak hosszú távon értelmezhető
  - Informális, szubjektív index helyett objektív, döntési szempontú („viselkedés, nem következtetés”) megközelítés
  - Elfogadási és elutasítási tartomány, *mindegy* hogy azon belül hová esik az empirikus tesztstatisztika (nincs bizonyíték-erősség)
  - A határt eredetileg még úgy gondolták, hogy a kétféle hiba súlya alapján határozzák meg a kutatók
  - „Hypothesis testing” (*nem* „significance testing”)

# Két iskola ütközése: J. Neyman és E. Pearson

- Döntés: „hibavalószínűség”
- *Két hipotézis között döntünk*
- Kísérletek sorát tételezzük fel...
- ...és a célunk, hogy *hosszú távon* biztosítsunk adott első- és másodfajú hibaarányt
- Egy adott, konkrét hipotézisről úgysem dönthető el, hogy mennyire igaz, a kérdés csak *hosszú távon* értelmezhető
- Informális, szubjektív index helyett objektív, döntési szempontú („viselkedés, nem következtetés”) megközelítés
- Elfogadási és elutasítási tartomány, *mindegy* hogy azon belül hová esik az empirikus tesztstatisztika (nincs bizonyíték-erősség)
- A határt eredetileg még úgy gondolták, hogy a kétféle hiba súlya alapján határozzák meg a kutatók
- „Hypothesis testing” (*nem* „significance testing”)

# Két iskola ütközése: J. Neyman és E. Pearson

- Döntés: „hibavalószínűség”
- Két hipotézis között döntünk
- Kísérletek sorát tételezzük fel...
- ...és a célunk, hogy *hosszú távon* biztosítsunk adott első- és másodfajú hibaarányt
- Egy adott, konkrét hipotézisről úgysem dönthető el, hogy mennyire igaz, a kérdés csak *hosszú távon* értelmezhető
- Informális, szubjektív index helyett objektív, döntési szempontú („viselkedés, nem következtetés”) megközelítés
- Elfogadási és elutasítási tartomány, *mindegy* hogy azon belül hová esik az empirikus tesztstatisztika (nincs bizonyíték-erősség)
- A határt eredetileg még úgy gondolták, hogy a kétféle hiba súlya alapján határozzák meg a kutatók
- „Hypothesis testing” (*nem* „significance testing”)

# Két iskola ütközése: J. Neyman és E. Pearson

- Döntés: „hibavalószínűség”
- Két hipotézis között döntünk
- Kísérletek sorát tételezzük fel...
- ...és a célunk, hogy *hosszú távon* biztosítsunk adott első- és másodfajú hibaarányt
- Egy adott, konkrét hipotézisről úgysem dönthető el, hogy mennyire igaz, a kérdés csak hosszú távon értelmezhető
- Informális, szubjektív index helyett objektív, döntési szempontú („viselkedés, nem következtetés”) megközelítés
- Elfogadási és elutasítási tartomány, *mindegy* hogy azon belül hová esik az empirikus tesztstatisztika (nincs bizonyíték-erősség)
- A határt eredetileg még úgy gondolták, hogy a kétféle hiba súlya alapján határozzák meg a kutatók
- „Hypothesis testing” (*nem* „significance testing”)



# Két iskola ütközése: J. Neyman és E. Pearson

- Döntés: „hibavalószínűség”
- Két hipotézis között döntünk
- Kísérletek sorát tételezzük fel...
- ...és a célunk, hogy *hosszú távon* biztosítsunk adott első- és másodfajú hibaarányt
- Egy adott, konkrét hipotézisről úgysem dönthető el, hogy mennyire igaz, a kérdés csak hosszú távon értelmezhető
- Informális, szubjektív index helyett objektív, döntési szempontú („viselkedés, nem következtetés”) megközelítés
- Elfogadási és elutasítási tartomány, *mindegy* hogy azon belül hová esik az empirikus tesztstatisztika (nincs bizonyíték-erősség)
- A határt eredetileg még úgy gondolták, hogy a kétféle hiba súlya alapján határozzák meg a kutatók
- „Hypothesis testing” (*nem* „significance testing”)

## Két iskola ütközése: J. Neyman és E. Pearson

- Döntés: „hibavalószínűség”
- Két hipotézis között döntünk
- Kísérletek sorát tételezzük fel...
- ...és a célunk, hogy *hosszú távon* biztosítsunk adott első- és másodfajú hibaarányt
- Egy adott, konkrét hipotézisről úgysem dönthető el, hogy mennyire igaz, a kérdés csak hosszú távon értelmezhető
- Informális, szubjektív index helyett objektív, döntési szempontú („viselkedés, nem következtetés”) megközelítés
- Elfogadási és elutasítási tartomány, *mindegy* hogy azon belül hová esik az empirikus tesztstatisztika (nincs bizonyíték-erősség)
- A határt eredetileg még úgy gondolták, hogy a kétféle hiba súlya alapján határozzák meg a kutatók
- „Hypothesis testing” (*nem* „significance testing”)

## Két iskola ütközése: J. Neyman és E. Pearson

- Döntés: „hibavalószínűség”
- Két hipotézis között döntünk
- Kísérletek sorát tételezzük fel...
- ...és a célunk, hogy *hosszú távon* biztosítsunk adott első- és másodfajú hibaarányt
- Egy adott, konkrét hipotézisről úgysem dönthető el, hogy mennyire igaz, a kérdés csak hosszú távon értelmezhető
- Informális, szubjektív index helyett objektív, döntési szempontú („viselkedés, nem következtetés”) megközelítés
- Elfogadási és elutasítási tartomány, *mindegy* hogy azon belül hová esik az empirikus tesztstatisztika (nincs bizonyíték-erősség)
- A határt eredetileg még úgy gondolták, hogy a kétféle hiba súlya alapján határozzák meg a kutatók
- „Hypothesis testing” (*nem* „significance testing”)

## Két iskola ütközése: J. Neyman és E. Pearson

- Döntés: „hibavalószínűség”
- Két hipotézis között döntünk
- Kísérletek sorát tételezzük fel...
- ...és a célunk, hogy *hosszú távon* biztosítsunk adott első- és másodfajú hibaarányt
- Egy adott, konkrét hipotézisről úgysem dönthető el, hogy mennyire igaz, a kérdés csak hosszú távon értelmezhető
- Informális, szubjektív index helyett objektív, döntési szempontú („viselkedés, nem következtetés”) megközelítés
- Elfogadási és elutasítási tartomány, *mindegy* hogy azon belül hová esik az empirikus tesztstatisztika (nincs bizonyíték-erősség)
- A határt eredetileg még úgy gondolták, hogy a kétféle hiba súlya alapján határozzák meg a kutatók
- „Hypothesis testing” (*nem* „significance testing”)

## Két iskola ütközése: J. Neyman és E. Pearson

- Döntés: „hibavalószínűség”
- Két hipotézis között döntünk
- Kísérletek sorát tételezzük fel...
- ...és a célunk, hogy *hosszú távon* biztosítsunk adott első- és másodfajú hibaarányt
- Egy adott, konkrét hipotézisről úgysem dönthető el, hogy mennyire igaz, a kérdés csak hosszú távon értelmezhető
- Informális, szubjektív index helyett objektív, döntési szempontú („viselkedés, nem következtetés”) megközelítés
- Elfogadási és elutasítási tartomány, *mindegy* hogy azon belül hová esik az empirikus tesztstatisztika (nincs bizonyíték-erősség)
- A határt eredetileg még úgy gondolták, hogy a kétféle hiba súlya alapján határozzák meg a kutatók
- „Hypothesis testing” (*nem* „significance testing”)

# A két iskola inkonzisztenciája

- Döntési határ van (és mindegy, hogy hová esik az empirikus tesztstatisztika) vagy nincs határ, és egy folytonos mértékünk van a bizonyíték erősségére?
- „Az érdekel minket, hogy az adott gyanúsított ártatlan-e, vagy az, hogy hosszú távon csak kevés ártatlant küldünk börtönbe?”
- A legfontosabb baj: nem lehet egyszerre beállítani a hosszútávú hibavalószínűséget és egyúttal az egyedi eredmény bizonyítóerejéről is nyilatkozni!
- A jelenlegi orvosi statisztikai gyakorlat azért keveri össze – teljesen inkonzisztensen – a kettőt, hogy megteremtse ennek az *illúzióját*

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med.* 1999 Jun 15;130(12):995-1004. Goodman SN. p values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. *Am J Epidemiol.* 1993 Mar 1;137(5):485-96; discussion 497-501. Gigerenzer G. The superego, the ego, and the id in statistical reasoning. In: *A handbook for data analysis in the behavioral sciences: Methodological issues.* 1993:311-339. Lehmann EL. The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?. In: *Selected Works of EL Lehmann.* 2012:201-208.

# A két iskola inkonzisztenciája

- Döntési határ van (és mindegy, hogy hová esik az empirikus tesztstatisztika) vagy nincs határ, és egy folytonos mértékünk van a bizonyíték erősségére?
- „Az érdekel minket, hogy az adott gyanúsított ártatlan-e, vagy az, hogy hosszú távon csak kevés ártatlant küldünk börtönbe?”
- A legfontosabb baj: nem lehet egyszerre beállítani a hosszútávú hibavalószínűséget és egyúttal az egyedi eredmény bizonyítóerejéről is nyilatkozni!
- A jelenlegi orvosi statisztikai gyakorlat azért keveri össze – teljesen inkonzisztensen – a kettőt, hogy megteremtse ennek az *illúzióját*

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med.* 1999 Jun 15;130(12):995-1004. Goodman SN. p values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. *Am J Epidemiol.* 1993 Mar 1;137(5):485-96; discussion 497-501. Gigerenzer G. The superego, the ego, and the id in statistical reasoning. In: *A handbook for data analysis in the behavioral sciences: Methodological issues.* 1993:311-339. Lehmann EL. The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?. In: *Selected Works of EL Lehmann.* 2012:201-208.

# A két iskola inkonzisztenciája

- Döntési határ van (és mindegy, hogy hová esik az empirikus tesztstatisztika) vagy nincs határ, és egy folytonos mértékünk van a bizonyíték erősségére?
- „Az érdekel minket, hogy az adott gyanúsított ártatlan-e, vagy az, hogy hosszú távon csak kevés ártatlant küldünk börtönbe?”
- A legfontosabb baj: nem lehet egyszerre beállítani a hosszútávú hibavalószínűséget és egyúttal az egyedi eredmény bizonyítóerejéről is nyilatkozni!
- A jelenlegi orvosi statisztikai gyakorlat azért keveri össze – teljesen inkonzisztensen – a kettőt, hogy megteremtse ennek az *illúzióját*

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med.* 1999 Jun 15;130(12):995-1004. Goodman SN. p values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. *Am J Epidemiol.* 1993 Mar 1;137(5):485-96; discussion 497-501. Gigerenzer G. The superego, the ego, and the id in statistical reasoning. In: *A handbook for data analysis in the behavioral sciences: Methodological issues.* 1993:311-339. Lehmann EL. The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?. In: *Selected Works of EL Lehmann.* 2012:201-208.



# A két iskola inkonzisztenciája

- Döntési határ van (és mindegy, hogy hová esik az empirikus tesztstatisztika) vagy nincs határ, és egy folytonos mértékünk van a bizonyíték erősségére?
- „Az érdekel minket, hogy az adott gyanúsított ártatlan-e, vagy az, hogy hosszú távon csak kevés ártatlant küldünk börtönbe?”
- A legfontosabb baj: nem lehet egyszerre beállítani a hosszútávú hibavalószínűséget és egyúttal az egyedi eredmény bizonyítóerejéről is nyilatkozni!
- A jelenlegi orvosi statisztikai gyakorlat azért keveri össze – teljesen inkonzisztensen – a kettőt, hogy megteremtse ennek az *illúzióját*

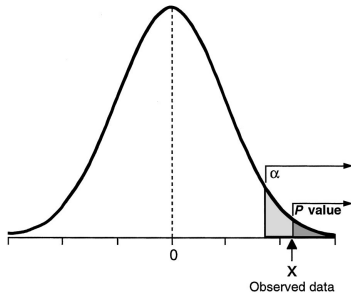
Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med.* 1999 Jun 15;130(12):995-1004. Goodman SN. p values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. *Am J Epidemiol.* 1993 Mar 1;137(5):485-96; discussion 497-501. Gigerenzer G. The superego, the ego, and the id in statistical reasoning. In: *A handbook for data analysis in the behavioral sciences: Methodological issues.* 1993:311-339. Lehmann EL. The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?. In: *Selected Works of EL Lehmann.* 2012:201-208.

# A két iskola inkonzisztenciája

- Döntési határ van (és mindegy, hogy hová esik az empirikus tesztstatisztika) vagy nincs határ, és egy folytonos mértékünk van a bizonyíték erősségére?
- „Az érdekel minket, hogy az adott gyanúsított ártatlan-e, vagy az, hogy hosszú távon csak kevés ártatlant küldünk börtönbe?”
- A legfontosabb baj: nem lehet egyszerre beállítani a hosszútávú hibavalószínűséget és egyúttal az egyedi eredmény bizonyítóerejéről is nyilatkozni!
- A jelenlegi orvosi statisztikai gyakorlat azért keveri össze – teljesen inkonzisztensen – a kettőt, hogy megteremtse ennek az *illúzióját*

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004. Goodman SN. p values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. Am J Epidemiol. 1993 Mar 1;137(5):485-96; discussion 497-501. Gigerenzer G. The superego, the ego, and the id in statistical reasoning. In: A handbook for data analysis in the behavioral sciences: Methodological issues. 1993:311-339. Lehmann EL. The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?. In: Selected Works of EL Lehmann. 2012:201-208.

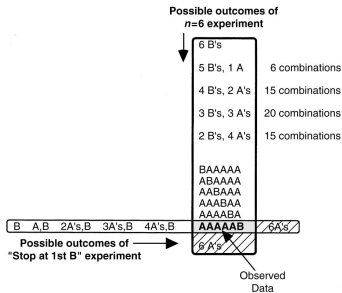
# A kompromisszum elkerülésének illúziója



Úgy tűnhet, hogy a  $p$ -érték is egyfajta hibavalószínűség  
„15. legjobb vagyok az évfolyamban”

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

## A kompromisszum elkerülésének illúziója



## Hosszútáv vs. rövidtáv

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med.* 1999 Jun 15;130(12):995-1004.