

# A legfontosabb félreértések a p-érték kapcsán

Ferenci Tamás  
tamas.ferenci@medstat.hu

2018. március 29.

# A legfontosabb félreértés a $p$ -érték kapcsán

- Az első és legfontosabb félreértés: a  $p$ -érték *csak és kizárólag* a mintavételi hibát jellemzi, a világon semmit, de semmit nem mond a nem-mintavételi hibákról
- ... miközben lehet, hogy egy kutatásban a nem-mintavételi hibák dominálnak!
- (Ebben az esetben a  $p$ -értékek közlése akár kifejezetten félrevezető is lehet!)
- Közel nem igaz tehát, hogy a  $p$ -érték (és társai) valamiféle univerzális metrikái lennének egy kutatás potenciális hibázásának

# A legfontosabb félreértés a $p$ -érték kapcsán

- Az első és legfontosabb félreértés: a  $p$ -érték *csak és kizárólag* a mintavételi hibát jellemzi, a világon semmit, de semmit nem mond a nem-mintavételi hibákról
- ... miközben lehet, hogy egy kutatásban a nem-mintavételi hibák dominálnak!
  - (Ebben az esetben a  $p$ -értékek közlése akár kifejezetten félrevezető is lehet!)
  - Közel nem igaz tehát, hogy a  $p$ -érték (és társai) valamiféle univerzális metrikái lennének egy kutatás potenciális hibázásának

# A legfontosabb félreértés a $p$ -érték kapcsán

- Az első és legfontosabb félreértés: a  $p$ -érték *csak és kizárólag* a mintavételi hibát jellemzi, a világon semmit, de semmit nem mond a nem-mintavételi hibákról
- ... miközben lehet, hogy egy kutatásban a nem-mintavételi hibák dominálnak!
- (Ebben az esetben a  $p$ -értékek közlése akár kifejezetten félrevezető is lehet!)
- Közel nem igaz tehát, hogy a  $p$ -érték (és társai) valamiféle univerzális metrikái lennének egy kutatás potenciális hibázásának

# A legfontosabb félreértés a $p$ -érték kapcsán

- Az első és legfontosabb félreértés: a  $p$ -érték *csak és kizárólag* a mintavételi hibát jellemzi, a világon semmit, de semmit nem mond a nem-mintavételi hibákról
- ... miközben lehet, hogy egy kutatásban a nem-mintavételi hibák dominálnak!
- (Ebben az esetben a  $p$ -értékek közlése akár kifejezetten félrevezető is lehet!)
- Közel nem igaz tehát, hogy a  $p$ -érték (és társai) valamiféle univerzális metrikái lennének egy kutatás potenciális hibázásának

## Második legfontosabb félreértés: egy kis motiváció

- Terroristák a városban...
- Mellesleg mitől medikális ez a példa?
- ... ritka betegségek szűrése!
- Na és mi köze a  $p$ -értékekhez?
- Íme...

## Második legfontosabb félreértés: egy kis motiváció

- Terroristák a városban...
- Mellesleg mitől medikális ez a példa?
- ... ritka betegségek szűrése!
- Na és mi köze a  $p$ -értékekhez?
- Íme...

## Második legfontosabb félreértés: egy kis motiváció

- Terroristák a városban...
- Mellesleg mitől medikális ez a példa?
- ...ritka betegségek szűrése!
  - Na és mi köze a  $p$ -értékekhez?
  - Íme...



## Második legfontosabb félreértés: egy kis motiváció

- Terroristák a városban...
- Mellesleg mitől medikális ez a példa?
- ... ritka betegségek szűrése!
- Na és mi köze a  $p$ -értékekhez?
- Íme...

## Második legfontosabb félreértés: egy kis motiváció

- Terroristák a városban...
- Mellesleg mitől medikális ez a példa?
- ... ritka betegségek szűrése!
- Na és mi köze a  $p$ -értékekhez?
- Íme...

## A második legfontosabb félreértés: $p$ -érték = hibavalószínűsége

- Az egyszerűség kedvéért a kísérletünknek csak két kimenete lehet: a gyógyszer hatásos és hatástalan
- Kétféle hiba van, legyen a hatástalan hatásosnak minősítése 5, a hatásos hatástalannak minősítése 20% valószínűségű
- Az utóbbi azt jelenti, hogy 80% az erő...
- ... az előbbi épp a választott szignifikanciaszint
- A hatástalanság a priori valószínűsége 90%
- Hogy mi? Ez miért érdekes? Hogy jön ez ide?

## A második legfontosabb félreértés: $p$ -érték = hibavalószínűsége

- Az egyszerűség kedvéért a kísérletünknek csak két kimenete lehet: a gyógyszer hatásos és hatástalan
- Kétféle hiba van, legyen a hatástalan hatásosnak minősítése 5, a hatásos hatástalannak minősítése 20% valószínűségű
- Az utóbbi azt jelenti, hogy 80% az erő...
- ... az előbbi épp a választott szignifikanciaszint
- A hatástalanság a priori valószínűsége 90%
- Hogy mi? Ez miért érdekes? Hogy jön ez ide?

## A második legfontosabb félreértés: $p$ -érték = hibavalószínűsége

- Az egyszerűség kedvéért a kísérletünknek csak két kimenete lehet: a gyógyszer hatásos és hatástalan
- Kétféle hiba van, legyen a hatástalan hatásosnak minősítése 5, a hatásos hatástalannak minősítése 20% valószínűségű
- Az utóbbi azt jelenti, hogy 80% az erő...
- ... az előbbi épp a választott szignifikanciaszint
- A hatástalanság a priori valószínűsége 90%
- Hogy mi? Ez miért érdekes? Hogy jön ez ide?

## A második legfontosabb félreértés: $p$ -érték = hibavalószínűsége

- Az egyszerűség kedvéért a kísérletünknek csak két kimenete lehet: a gyógyszer hatásos és hatástalan
- Kétféle hiba van, legyen a hatástalan hatásosnak minősítése 5, a hatásos hatástalannak minősítése 20% valószínűségű
- Az utóbbi azt jelenti, hogy 80% az erő. . .
- . . . az előbbi épp a választott szignifikanciaszint
- A hatástalanság a priori valószínűsége 90%
- Hogy mi? Ez miért érdekes? Hogy jön ez ide?

## A második legfontosabb félreértés: $p$ -érték = hibavalószínűsége

- Az egyszerűség kedvéért a kísérletünknek csak két kimenete lehet: a gyógyszer hatásos és hatástalan
- Kétféle hiba van, legyen a hatástalan hatásosnak minősítése 5, a hatásos hatástalannak minősítése 20% valószínűségű
- Az utóbbi azt jelenti, hogy 80% az erő. . .
- . . . az előbbi épp a választott szignifikanciaszint
- A hatástalanság a priori valószínűsége 90%
- Hogy mi? Ez miért érdekes? Hogy jön ez ide?

## A második legfontosabb félreértés: $p$ -érték = hibavalószínűsége

- Az egyszerűség kedvéért a kísérletünknek csak két kimenete lehet: a gyógyszer hatásos és hatástalan
- Kétféle hiba van, legyen a hatástalan hatásosnak minősítése 5, a hatásos hatástalannak minősítése 20% valószínűségű
- Az utóbbi azt jelenti, hogy 80% az erő. . .
- . . . az előbbi épp a választott szignifikanciaszint
- A hatástalanság a priori valószínűsége 90%
- Hogy mi? Ez miért érdekes? Hogy jön ez ide?



# Egy kis fejszámítás

- Hogy matematikai részletek nélkül megindokoljam ezt, végezzünk egy gondolatkísérletet
- 1000-szer szimuláljuk ezt a világot (avagy 1000 párhuzamos univerzumot tekintünk)
- $1000 \cdot 0,1 = 100$ -szor hatásos lesz a gyógyszer. . .  
1000-szor szimuláljuk ezt a világot, azaz 1000 párhuzamos univerzumot tekintünk.
- $1000 \cdot 0,9 = 900$ -szor hatástalan lesz a gyógyszer. . .  
1000-szor szimuláljuk ezt a világot, azaz 1000 párhuzamos univerzumot tekintünk.
- Összességében  $80 + 45 = 125$  esetben lesz hatásosnak minősítve a gyógyszerünk
- A hibarány, tehát, hogy a hatásosnak minősítések közül mekkora arányban hatástalan valójában a gyógyszer  $45/125 = 36\%$  – *marhára nem 5%!*

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci. 2014 Nov 19;1(3):140216.

# Egy kis fejszámítás

- Hogy matematikai részletek nélkül megindokoljam ezt, végezzünk egy gondolatkísérletet
- 1000-szer szimuláljuk ezt a világot (avagy 1000 párhuzamos univerzumot tekintünk)
- $1000 \cdot 0,1 = 100$ -szor hatásos lesz a gyógyszer. . .
  - Ebből  $100 \cdot 0,2 = 20$ -szor hatástalannak minősítjük, a maradék 80 esetben hatásosnak
- $1000 \cdot 0,9 = 900$ -szor hatástalan lesz a gyógyszer. . .
  - Ebből  $900 \cdot 0,1 = 90$ -szor hatástalannak minősítjük, a maradék 810 esetben hatásosnak
- Összességében  $80 + 45 = 125$  esetben lesz hatásosnak minősítve a gyógyszerünk
- A hibarány, tehát, hogy a hatásosnak minősítések közül mekkora arányban hatástalan valójában a gyógyszer  $45/125 = 36\%$  – *marhára nem 5%!*

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci. 2014 Nov 19;1(3):140216.

# Egy kis fejszámítás

- Hogy matematikai részletek nélkül megindokoljam ezt, végezzünk egy gondolatkísérletet
  - 1000-szer szimuláljuk ezt a világot (avagy 1000 párhuzamos univerzumot tekintünk)
  - $1000 \cdot 0,1 = 100$ -szor hatásos lesz a gyógyszer. . .
    - . . . ebből  $100 \cdot 0,2 = 20$ -szor hatástalannak minősítjük, a maradék 80 esetben hatásosnak
  - $1000 \cdot 0,9 = 900$ -szor hatástalan lesz a gyógyszer. . .
- Összesítés:
- Összességében  $80 + 45 = 125$  esetben lesz hatásosnak minősítve a gyógyszerünk
  - A hibarány, tehát, hogy a hatásosnak minősítések közül mekkora arányban hatástalan valójában a gyógyszer  $45/125 = 36\%$  – *marhára nem 5%!*

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci. 2014 Nov 19;1(3):140216.

# Egy kis fejszámítás

- Hogy matematikai részletek nélkül megindokoljam ezt, végezzünk egy gondolatkísérletet
- 1000-szer szimuláljuk ezt a világot (avagy 1000 párhuzamos univerzumot tekintünk)
- $1000 \cdot 0,1 = 100$ -szor hatásos lesz a gyógyszer...
  - ... ebből  $100 \cdot 0,2 = 20$ -szor hatástalannak minősítjük, a maradék 80 esetben hatásosnak
- $1000 \cdot 0,9 = 900$ -szor hatástalan lesz a gyógyszer...
  - ... ebből  $900 \cdot 0,05 = 45$ -szor hatásosnak minősítjük, a maradék 855 esetben hatástalannak
- Összességében  $80 + 45 = 125$  esetben lesz hatásosnak minősítve a gyógyszerünk
- A hibarány, tehát, hogy a hatásosnak minősítések közül mekkora arányban hatástalan valójában a gyógyszer  $45/125 = 36\%$  – *marhára nem 5%!*

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci. 2014 Nov 19;1(3):140216.

# Egy kis fejszámítás

- Hogy matematikai részletek nélkül megindokoljam ezt, végezzünk egy gondolatkísérletet
- 1000-szer szimuláljuk ezt a világot (avagy 1000 párhuzamos univerzumot tekintünk)
- $1000 \cdot 0,1 = 100$ -szor hatásos lesz a gyógyszer. . .
  - . . . ebből  $100 \cdot 0,2 = 20$ -szor hatástalannak minősítjük, a maradék 80 esetben hatásosnak
- $1000 \cdot 0,9 = 900$ -szor hatástalan lesz a gyógyszer. . .
  - . . . ebből  $900 \cdot 0,05 = 45$ -szor hatásosnak minősítjük, a maradék 855 esetben hatástalannak
- Összességében  $80 + 45 = 125$  esetben lesz hatásosnak minősítve a gyógyszerünk
- A hibarány, tehát, hogy a hatásosnak minősítések közül mekkora arányban hatástalan valójában a gyógyszer  $45/125 = 36\%$  – *marhára nem 5%!*

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci. 2014 Nov 19;1(3):140216.

# Egy kis fejszámítás

- Hogy matematikai részletek nélkül megindokoljam ezt, végezzünk egy gondolatkísérletet
- 1000-szer szimuláljuk ezt a világot (avagy 1000 párhuzamos univerzumot tekintünk)
- $1000 \cdot 0,1 = 100$ -szor hatásos lesz a gyógyszer. . .
  - . . . ebből  $100 \cdot 0,2 = 20$ -szor hatástalannak minősítjük, a maradék 80 esetben hatásosnak
- $1000 \cdot 0,9 = 900$ -szor hatástalan lesz a gyógyszer. . .
  - . . . ebből  $900 \cdot 0,05 = 45$ -szor hatásosnak minősítjük, a maradék 855 esetben hatástalannak
- Összességében  $80 + 45 = 125$  esetben lesz hatásosnak minősítve a gyógyszerünk
- A hibarány, tehát, hogy a hatásosnak minősítések közül mekkora arányban hatástalan valójában a gyógyszer  $45/125 = 36\%$  – *marhára nem 5%!*

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci. 2014 Nov 19;1(3):140216.

# Egy kis fejszámítás

- Hogy matematikai részletek nélkül megindokoljam ezt, végezzünk egy gondolatkísérletet
- 1000-szer szimuláljuk ezt a világot (avagy 1000 párhuzamos univerzumot tekintünk)
- $1000 \cdot 0,1 = 100$ -szor hatásos lesz a gyógyszer. . .
  - . . . ebből  $100 \cdot 0,2 = 20$ -szor hatástalannak minősítjük, a maradék 80 esetben hatásosnak
- $1000 \cdot 0,9 = 900$ -szor hatástalan lesz a gyógyszer. . .
  - . . . ebből  $900 \cdot 0,05 = 45$ -szor hatásosnak minősítjük, a maradék 855 esetben hatástalannak
- Összességében  $80 + 45 = 125$  esetben lesz hatásosnak minősítve a gyógyszerünk
- A hibarány, tehát, hogy a hatásosnak minősítések közül mekkora arányban hatástalan valójában a gyógyszer  $45/125 = 36\%$  – *marhára nem 5%!*

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci. 2014 Nov 19:1(3):140216.

# Egy kis fejszámítás

- Hogy matematikai részletek nélkül megindokoljam ezt, végezzünk egy gondolat kísérletet
- 1000-szer szimuláljuk ezt a világot (avagy 1000 párhuzamos univerzumot tekintünk)
- $1000 \cdot 0,1 = 100$ -szor hatásos lesz a gyógyszer. . .
  - . . . ebből  $100 \cdot 0,2 = 20$ -szor hatástalannak minősítjük, a maradék 80 esetben hatásosnak
- $1000 \cdot 0,9 = 900$ -szor hatástalan lesz a gyógyszer. . .
  - . . . ebből  $900 \cdot 0,05 = 45$ -szor hatásosnak minősítjük, a maradék 855 esetben hatástalannak
- Összességében  $80 + 45 = 125$  esetben lesz hatásosnak minősítve a gyógyszerünk
- A hibarány, tehát, hogy a hatásosnak minősítések közül mekkora arányban hatástalan valójában a gyógyszer  $45/125 = 36\%$  – *marhára nem 5%!*

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci. 2014 Nov 19;1(3):140216.



# Egy kis fejszámítás

- Hogy matematikai részletek nélkül megindokoljam ezt, végezzünk egy gondolatkísérletet
- 1000-szer szimuláljuk ezt a világot (avagy 1000 párhuzamos univerzumot tekintünk)
- $1000 \cdot 0,1 = 100$ -szor hatásos lesz a gyógyszer. . .
  - . . . ebből  $100 \cdot 0,2 = 20$ -szor hatástalannak minősítjük, a maradék 80 esetben hatásosnak
- $1000 \cdot 0,9 = 900$ -szor hatástalan lesz a gyógyszer. . .
  - . . . ebből  $900 \cdot 0,05 = 45$ -szor hatásosnak minősítjük, a maradék 855 esetben hatástalannak
- Összességében  $80 + 45 = 125$  esetben lesz hatásosnak minősítve a gyógyszerünk
- A hibarány, tehát, hogy a hatásosnak minősítések közül mekkora arányban hatástalan valójában a gyógyszer  $45/125 = 36\%$  – *marhára nem 5%!*

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci. 2014 Nov 19;1(3):140216.

# A hatás

Power of study (proportion (%) of time we reject null hypothesis if it is false)	Percentage of "significant" results that are false positives		
	P=0.05	P=0.01	P=0.001
<b>80% of ideas correct (null hypothesis false)</b>			
20	5.9	1.2	0.10
50	2.4	0.5	0.05
80	1.5	0.3	0.03
<b>50% of ideas correct (null hypothesis false)</b>			
20	20.0	4.8	0.50
50	9.1	2.0	0.20
80	5.9	1.2	0.10
<b>10% of ideas correct (null hypothesis false)</b>			
20	69.2	31.0	4.30
50	47.4*	15.3	1.80
80	36.0	10.1	1.10
<b>1% of ideas correct (null hypothesis false)</b>			
20	96.1	83.2	33.10
50	90.8	66.4	16.50
80	86.1	55.3	11.00

\*Corresponds to assumptions in table 2.

Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? BMJ. 2001 Jan 27;322(7280):226-31.

# A probléma oka

- A kutya ott van elásva, hogy mi a prior valószínűsége
- A  $p$ -érték „előnye”, hogy ez nem kell hozzá
- De emiatt nem is azt méri, amit sokan gondolnának!
- Azt is lehet, de ahhoz kell a prior valószínűség: Bayes-tétel

$$\mathbb{P}(H_0|\text{Minta}) = \frac{\mathbb{P}(\text{Minta}|H_0) \cdot \mathbb{P}(H_0)}{\mathbb{P}(\text{Minta})}$$

- Út a Bayes-faktorokhoz

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. Ann Intern Med. 1999 Jun 15;130(12):1005-13.

# A probléma oka

- A kutya ott van elásva, hogy mi a prior valószínűsége
- A  $p$ -érték „előnye”, hogy ez nem kell hozzá
- De emiatt nem is azt méri, amit sokan gondolnának!
- Azt is lehet, de ahhoz kell a prior valószínűség: Bayes-tétel

$$\mathbb{P}(H_0|\text{Minta}) = \frac{\mathbb{P}(\text{Minta}|H_0) \cdot \mathbb{P}(H_0)}{\mathbb{P}(\text{Minta})}$$

- Út a Bayes-faktorokhoz

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. Ann Intern Med. 1999 Jun 15;130(12):1005-13.

# A probléma oka

- A kutya ott van elásva, hogy mi a prior valószínűsége
- A  $p$ -érték „előnye”, hogy ez nem kell hozzá
- De emiatt nem is azt méri, amit sokan gondolnának!
- Azt is lehet, de ahhoz kell a prior valószínűség: Bayes-tétel

$$\mathbb{P}(H_0|\text{Minta}) = \frac{\mathbb{P}(\text{Minta}|H_0) \cdot \mathbb{P}(H_0)}{\mathbb{P}(\text{Minta})}$$

- Út a Bayes-faktorokhoz

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. Ann Intern Med. 1999 Jun 15;130(12):1005-13.

# A probléma oka

- A kutya ott van elásva, hogy mi a prior valószínűsége
- A  $p$ -érték „előnye”, hogy ez nem kell hozzá
- De emiatt nem is azt méri, amit sokan gondolnának!
- Azt is lehet, de ahhoz kell a prior valószínűség: Bayes-tétel

$$\mathbb{P}(H_0|\text{Minta}) = \frac{\mathbb{P}(\text{Minta}|H_0) \cdot \mathbb{P}(H_0)}{\mathbb{P}(\text{Minta})}$$

- Út a Bayes-faktorokhoz

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med.* 1999 Jun 15;130(12):995-1004. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med.* 1999 Jun 15;130(12):1005-13.

# A probléma oka

- A kutya ott van elásva, hogy mi a prior valószínűsége
- A  $p$ -érték „előnye”, hogy ez nem kell hozzá
- De emiatt nem is azt méri, amit sokan gondolnának!
- Azt is lehet, de ahhoz kell a prior valószínűség: Bayes-tétel

$$\mathbb{P}(H_0|\text{Minta}) = \frac{\mathbb{P}(\text{Minta}|H_0) \cdot \mathbb{P}(H_0)}{\mathbb{P}(\text{Minta})}$$

- Út a Bayes-faktorokhoz

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med.* 1999 Jun 15;130(12):995-1004. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med.* 1999 Jun 15;130(12):1005-13.

# A probléma oka

- A kutya ott van elásva, hogy mi a prior valószínűsége
- A  $p$ -érték „előnye”, hogy ez nem kell hozzá
- De emiatt nem is azt méri, amit sokan gondolnának!
- Azt is lehet, de ahhoz kell a prior valószínűség: Bayes-tétel

$$\mathbb{P}(H_0|\text{Minta}) = \frac{\mathbb{P}(\text{Minta}|H_0) \cdot \mathbb{P}(H_0)}{\mathbb{P}(\text{Minta})}$$

- Út a Bayes-faktorokhoz

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. Ann Intern Med. 1999 Jun 15;130(12):1005-13.



# A $p$ -értékek minősítése, a nagyságának a jelentősége

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE $P < 0.10$ LEVEL
0.08	
0.09	
0.099	HEY LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
$\geq 0.1$	

Aki azt gondolná, hogy ez egy vicc: <https://mchankins.wordpress.com/2013/04/21/still-not-significant-2/>