

# Ez volt a HF

HF1: Három közvélemény-kutató megmérte egy párt támogatottságát, az alábbi eredményekkel:

- 1) 36.2%, 600 fős minta
- 2) 29.8%, 2500 fős minta
- 3) 31.4%, 1000 fős minta

Ezek hibahatáron belüli eltérések?

Az eredmények összesítésével milyen eredményt kapunk? Milyen hibahatárral?

HF2: Hajtsuk végre a <https://ruzsaz.github.io/horoszkop.csv> elemzését R-ben, az alábbi lépésekkel:

- 1) Írunk függvényt, amely egy bemenő paraméter (1-12-ig) alapján kiszámolja az adott hónapban születettek élettartamát, illetve a 95%-os konfidenciaintervallumot.
- 2) Csinálunk egy 12 elemű vektort, ami 1-12-ig tartalmazza a számokat.
- 3) Hívjuk meg a függvényt a vektor elemeire.
- 4) Ábrázoljuk az eredményeket (becslést és konfidenciaintervallumot).



# Tematika

- ~~Leíró statisztika (átlag, median, módsz, szórás...)~~
- Paraméterbecslés (mennyi?)
- Hipotézisvizsgálat (igaz vagy nem?)
- Regressziószámítás, modellezés (hogy befolyásol?)





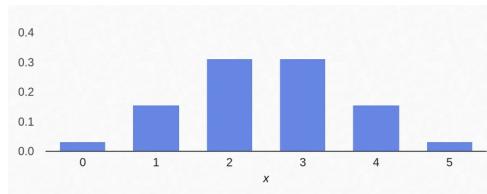
# Szabályos?

$$p = P(\text{fej}) = 0.5?$$

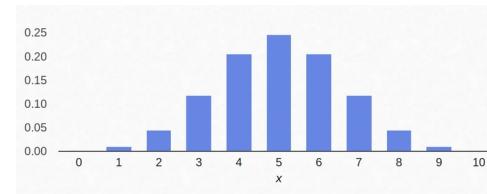


# Terv

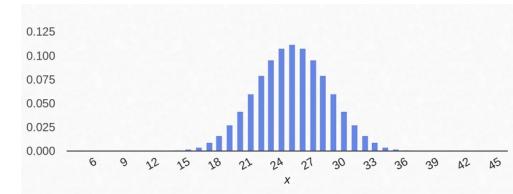
- 1) Tegyük fel, hogy  $p=0.5$ . ( $H_0$ , nullhipotézis)
- 2) Végezzünk  $n$  kísérletet. Ekkor a fejek eloszlása  $B(n, 0.5)$ .  
(Próbastatisztika eloszlása...)
- 3) Hihető, hogy az eredmény  $B(n, 0.5)$ -ből származik?



$B(5, 0.5)$



$B(10, 0.5)$



$B(50, 0.5)$



# Mi a következtetés?

$n$	fej	szabályos	hamis	passz
1	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
100	21	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
100	39	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
100	46	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



# Mi a következtetés? 95% konfidenciával

$n$	fej	szabályos	hamis	passz
1	0	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3	0	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
6	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6	2	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
10	2	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
10	5	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
100	21	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
100	39	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
100	46	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>



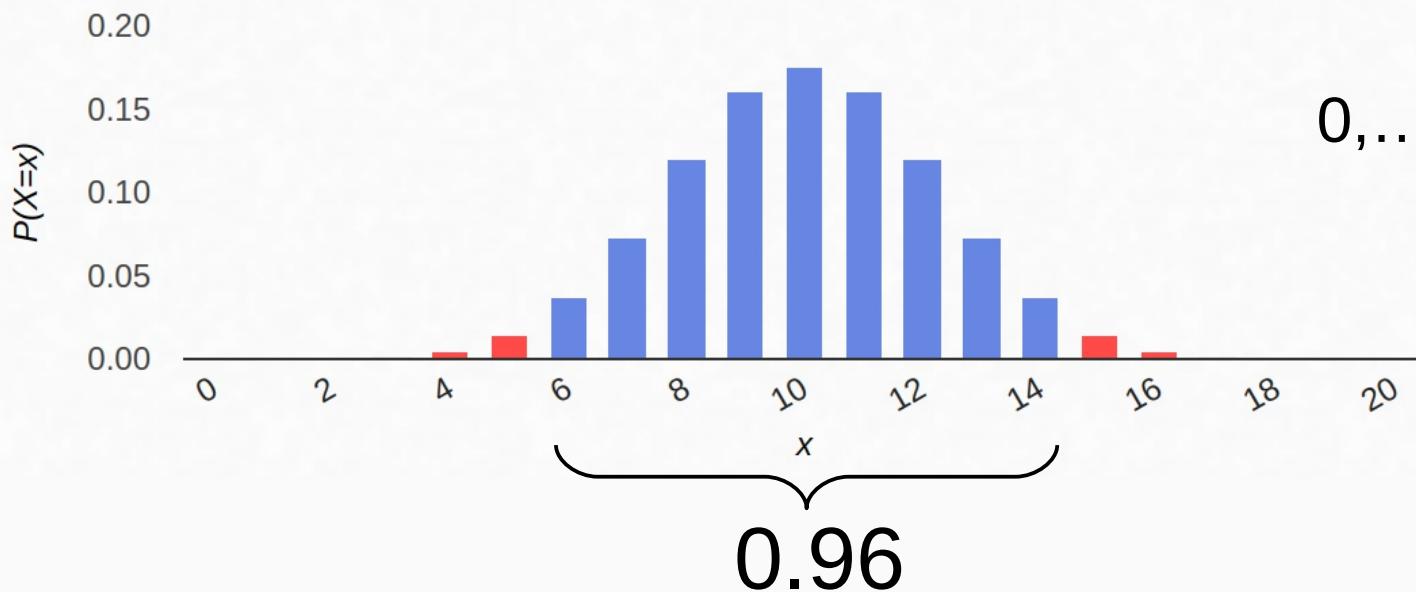
$B(20, 0.5)$ ,

95% konfidenciaszint

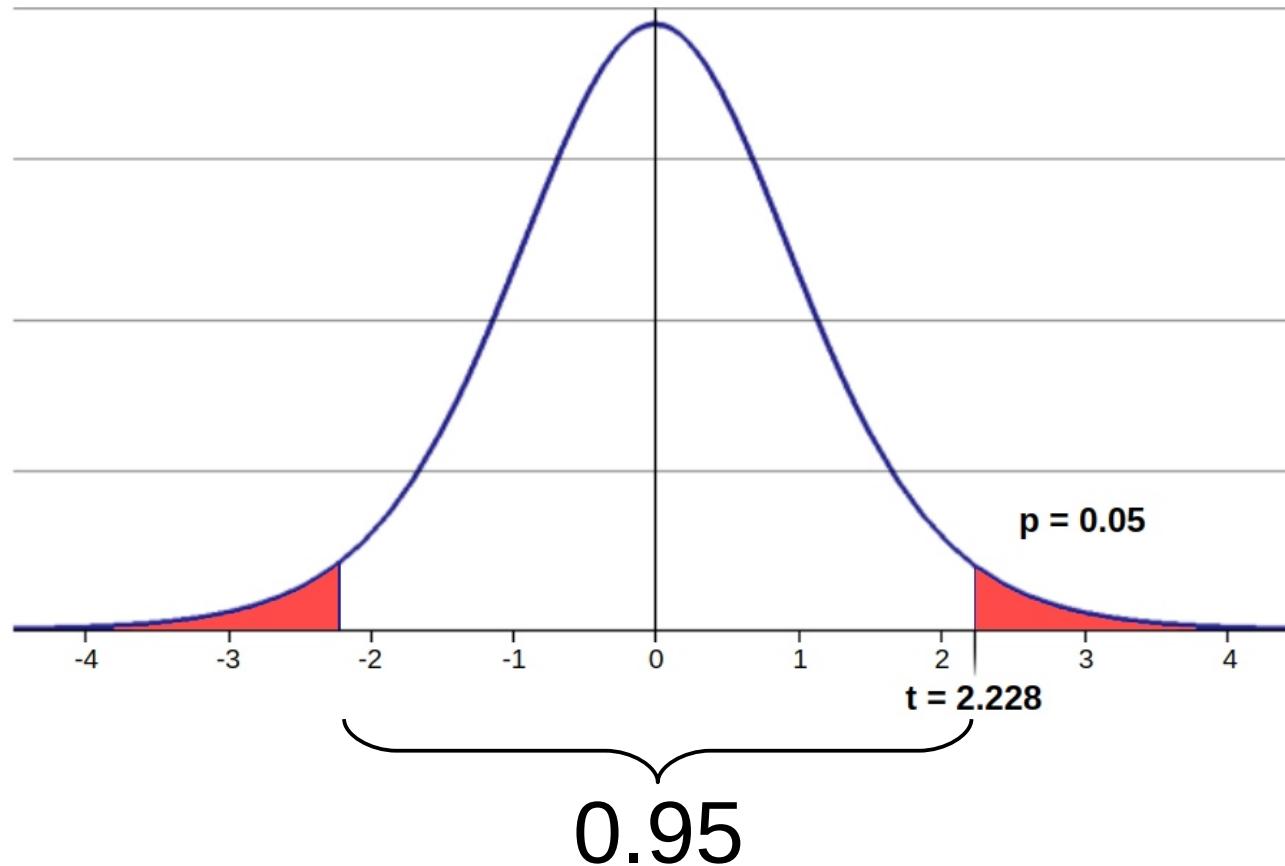
20 dobás, fejek száma

6, 7, ..., 14 hihető

0, ..., 5, 15, ..., 20 nem



# T-eloszlás (10 df), 95% konfidenciaszint



„Senki nem tekinthető bűnösnek mindaddig,  
amíg büntetőjogi felelősséget a bíróság  
jogerős határozata nem állapította meg.”



# Hipotézisvizsgálat

- 1) Feltesszük a nullhipotézist ( $H_0$ ): nincs semmi különös.
- 2) Kisérleteket végzünk.
- 3a) Ha  $H_0$  esetén a tapasztalt eredmény valószínűtlen, akkor elfogadjuk, hogy  $H_0$  nem igaz.
- 3b) Ha  $H_0$ -al összevágó eredményt tapasztalunk, akkor nem sikerült megcáfoltuk  $H_0$ -at. (De ettől még NEM BIZONYÍTOTTUK  $H_0$  fennállását.)

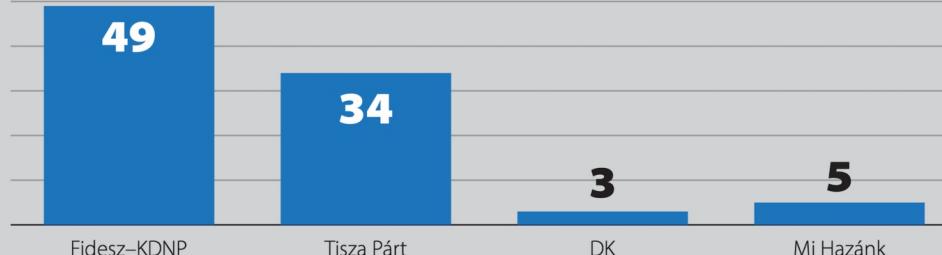
# Bírósági tárgyalás

- 1) A vádlott ártatlan, míg az ellenkezőjét nem bizonyítják.
- 2) Az ügyész ével a bűnösség mellett.
- 3a) Ha az érvek alapján nagyon valószínűtlen az ártatlanság, akkor bizonyítottnak tekintjük, hogy bűnös.
- 3b) Ha a bemutatott bizonyítékok mellett elképzelhető a vádlott ártatlansága, akkor nem ítélik bűnösnek. (De ettől még NEM BIZONYÍTOTT hogy ártatlan.)

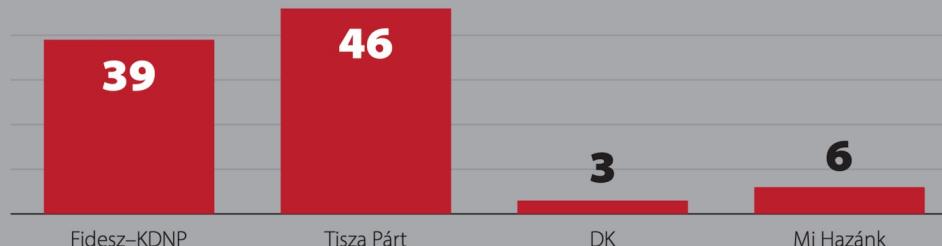


# Ki hazudik?

Legvalósínűbb listás eredmény a Nézőpont Intézetnél (százalék)



Választani tudó biztos szavazók a Mediánnál (százalék)



(demokrata.hu)

„Az Nézőpont Intézet kutatása 2024. október 14. és 17. között, 1000 fő telefonos megkérdezésével készült. A minta reprezentatív a 18 évesnél idősebb lakosságra nem, kor, régió, településtípus és iskolai végzettség szerint. 1000 fős mintanagyság és 95 százalékos megbízhatósági szint esetén a mintavételi hiba  $\pm 3,16$  százalék.” (Nézőpont intézet)

„Medián legfrissebb felmérése szerint, amit 1000 fős reprezentatív minta telefonos megkérdezésével készített október 28-a és november 4-e között a HVG számára.” (HVG)



# Milyen próbát alkalmazzunk?

- 1) 1 mintás? 2 mintás?
- 2) T-próba? Khi-négyzet? Z-próba? Binomiális? Anova?
- 3) Egyoldali? Két oldali?
- 4) Várható érték azonosságára? Előfordulás gyakoriságára?  
Szórásokra?





ace & Grains

Salad Dressing 4A

Crackers 4B

4A

# Z-test for comparing two proportions



## Comparing the proportions of two binomials [edit]

Main article: [Two-proportion Z-test](#)

The **z-test for comparing two proportions** is a statistical method used to evaluate whether the proportion of a certain characteristic differs significantly between two independent samples. This test leverages the property that the [sample proportions](#) (which is the average of observations coming from a [Bernoulli distribution](#)) are [asymptotically normal](#) under the [Central Limit Theorem](#), enabling the construction of a z-test.

The z-statistic for comparing two proportions is computed using:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Where:

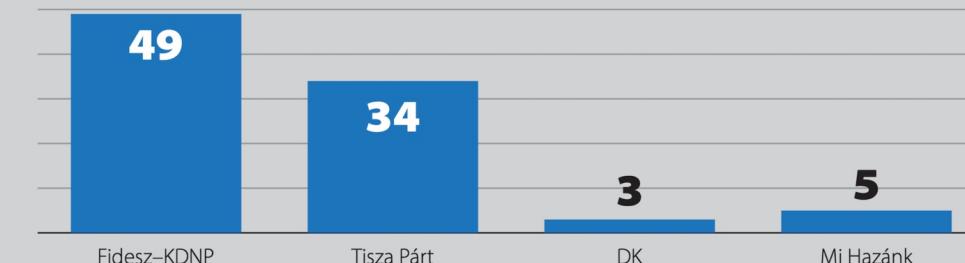
- $\hat{p}_1$  = sample proportion in the first sample
- $\hat{p}_2$  = sample proportion in the second sample
- $n_1$  = size of the first sample
- $n_2$  = size of the second sample
- $\hat{p}$  = pooled proportion, calculated as  $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$ , where  $x_1$  and  $x_2$  are the counts of successes in the two samples.

The [confidence interval](#) for the difference between two proportions, based on the definitions above, is:

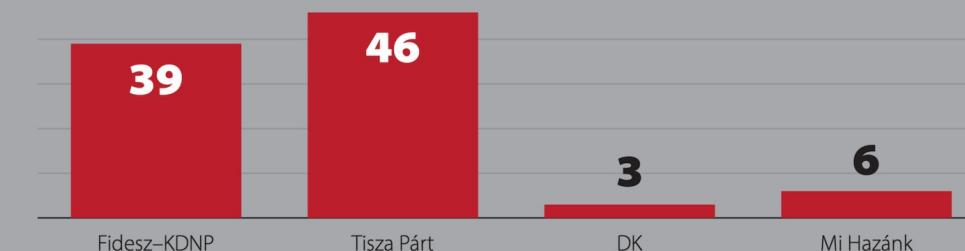
$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

# Feladat: Z-test for comparing two proportions

Legvalószínűbb listás eredmény a Nézőpont Intézetnél (százalék)



Választani tudó biztos szavazók a Mediánnál (százalék)



<https://colab.research.google.com/>

- 1) Hajtsuk végre a tesztet.
- 2) Értelmezzük az eredményt.
- 3) Ábrázoljuk.



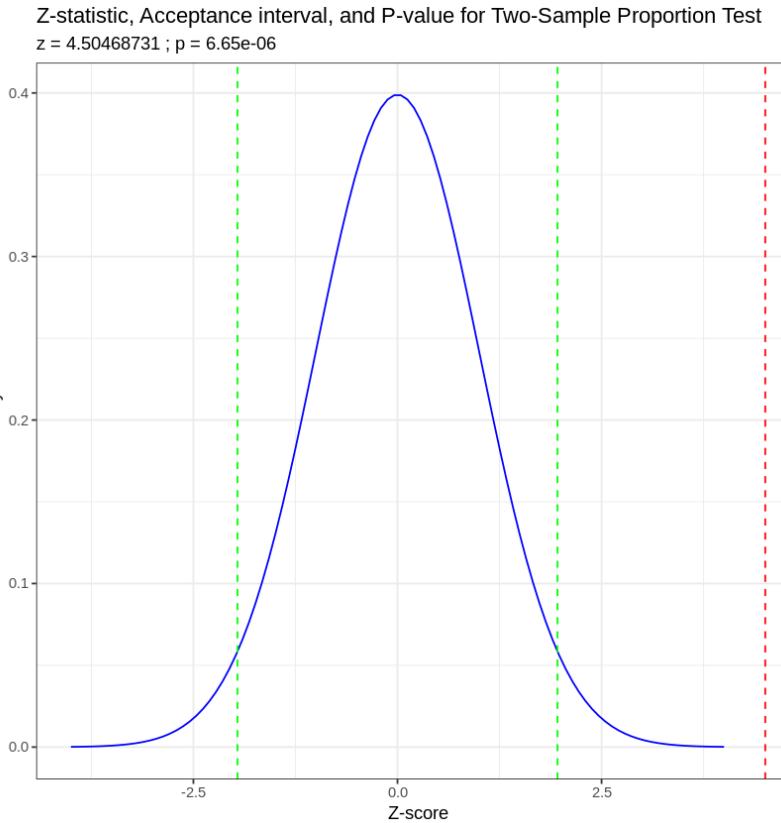
<https://colab.research.google.com/>

1) „make a z test for testing the equality of two proportions”

Z-statistic: 4.504687

P-value: 6.647074e-06

2) „draw the results, the z function, the 95% acceptance interval, and the z value from the previous calculation”



# HF1 <https://ruzsaz.github.io/stat2.pdf>

Egy homeopátiás altató működését teszteljük.

10 önkéntes elalvási idejét mértük 2 alkalommal, egyik alkalommal placebót kaptak, másik alkalommal az altatót. (Ők nem tudják, hogy mikor mit.)

Az elalváshoz szükséges idő (perc) így alakult:

<https://ruzsaz.github.io/homeopatia.csv>

	1	2	3	4	5	6	7	8	9	10
placebo	23.2	14	59.2	31.1	47.4	7.8	44.2	21.1	33.3	9.2
altató	31.5	7.8	54.9	33	11.1	23.3	43	16.7	12.6	8.6

Működik a szer?



HF2 <https://ruzsaz.github.io/stat2.pdf>

Szükség van egyáltalán hipotézisvizsgálatra???

Nem elég paraméterbecsléssel megbecsülni a két minta várható értékét, és megnézni, hogy hibahatárnál jobban különböznek-e?



# HF3 <https://ruzsaz.github.io/stat2.pdf>

2021-ben egy időszakban mindenkit beoltottak a COVID ellen. A lakók egy része visszautasította az előbb rendelkezésre álló kínai oltást, őket egy picit később a Pfizer–BioNTech Comirnaty nevű termékével oltották be. Amikor néhány hónap múlva az intézetben végigsöpört a járvány, feljegyezték, hogy hányan úszták meg tünetek nélkül, hány lakónál tudták a betegséget az intézetben kezelni, és hány lakó került kórházba.

	tünetmentes	enyhe	kórház
Sinopharm	87	44	14
Comirnaty	46	19	6

Az adatok alapján állíthatjuk-e, hogy a Comirnaty volt a jobb választás?

