

# A véletlen szerepéről

Ferenci Tamás

`tamas.ferenci@medstat.hu`

Utoljára frissítve: 2024. október 1.

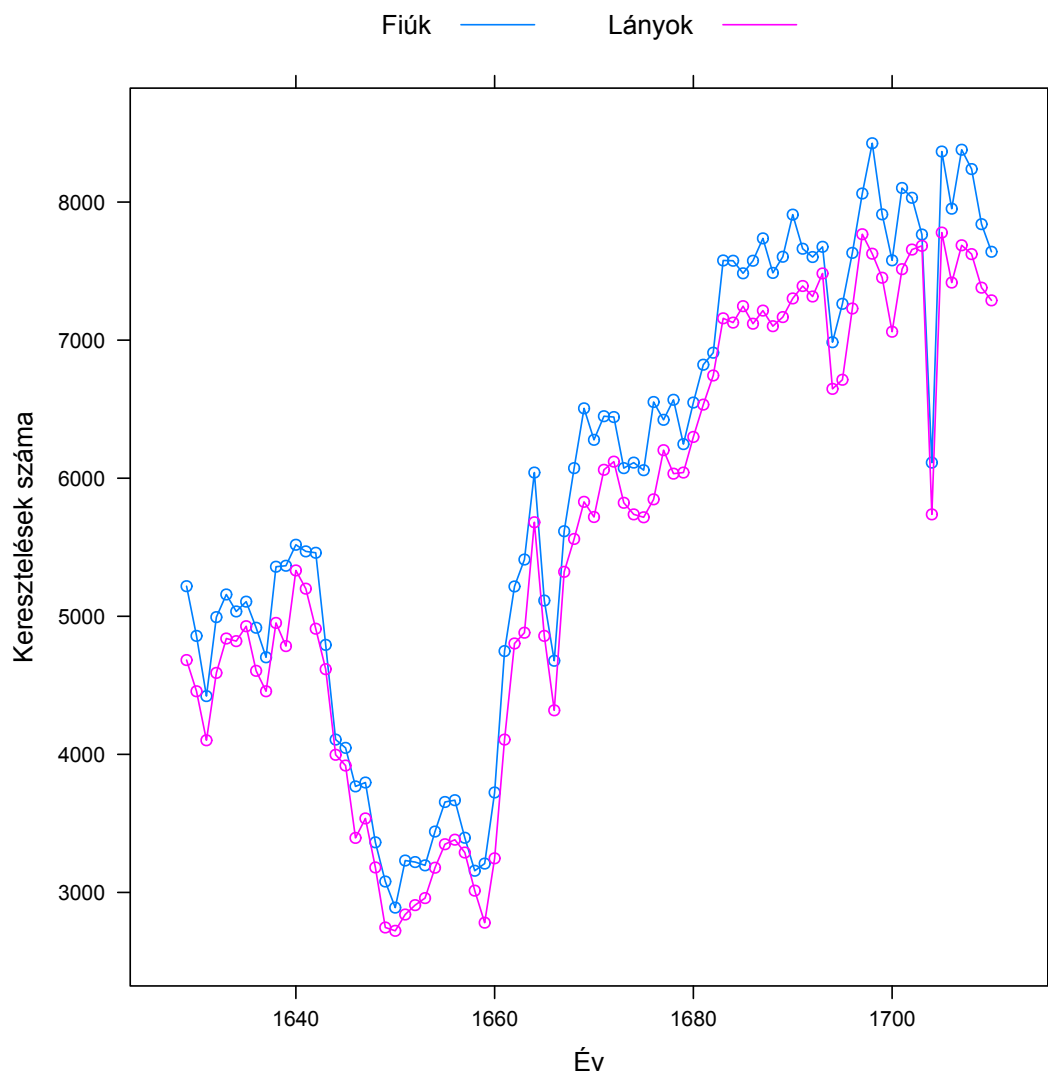
**Az első példa**



Az első példa

Christened.			Christened.		
Anno.	Males.	Females.	Anno.	Males.	Females.
1667	5616	5322	1689	7604	7167
68	6073	5560	90	7909	7302
69	6506	5829	91	7662	7392
70	6278	5719	92	7602	7316
71	6449	6061	93	7676	7483
72	6443	6120	94	6985	6647
73	6073	5822	95	7263	6713
74	6113	5738	96	7632	7229
75	6058	5717	97	8062	7767
76	6552	5847	98	8426	7626
77	6423	6203	99	7911	7452
78	6568	6033	1700	7578	7061
79	6247	6041	1701	8102	7514
80	6548	6299	1702	8031	7656
81	6822	6533	1703	7765	7683
82	6909	6744	1704	6113	5738
83	7577	7158	1705	8366	7779
84	7575	7127	1706	7952	7417
85	7484	7246	1707	8379	7687
86	7575	7119	1708	8239	7623
87	7737	7214	1709	7840	7380
88	7487	7101	1710	7640	7288

Az első példa



### Arbuthnot gondolatmenete

- Fiú/lány arány nem fele-fele: létezik Isten (teológiai kérdés, most elfogadjuk igaznak)
- De van egy statisztikai kérdés is: tényleg nem fele-fele?
- Analógia: minden év egy pénzfeldobás, fej – fiútöbbség, írás – lánytöbbség
- Arbuthnot okfejtése:
  1. Az, hogy 82-ből 82-ször fejet dobtunk *nem* jelenti *biztosan*, hogy a pénzérme nem szabályos (=létezik Isten), *mert* szabályos pénzérmével is lehet 82-ből 82-ször fejet dobni

2. *De*: ennek a valószínűsége extrém kicsi ( $2,1 \cdot 10^{-25}$ )

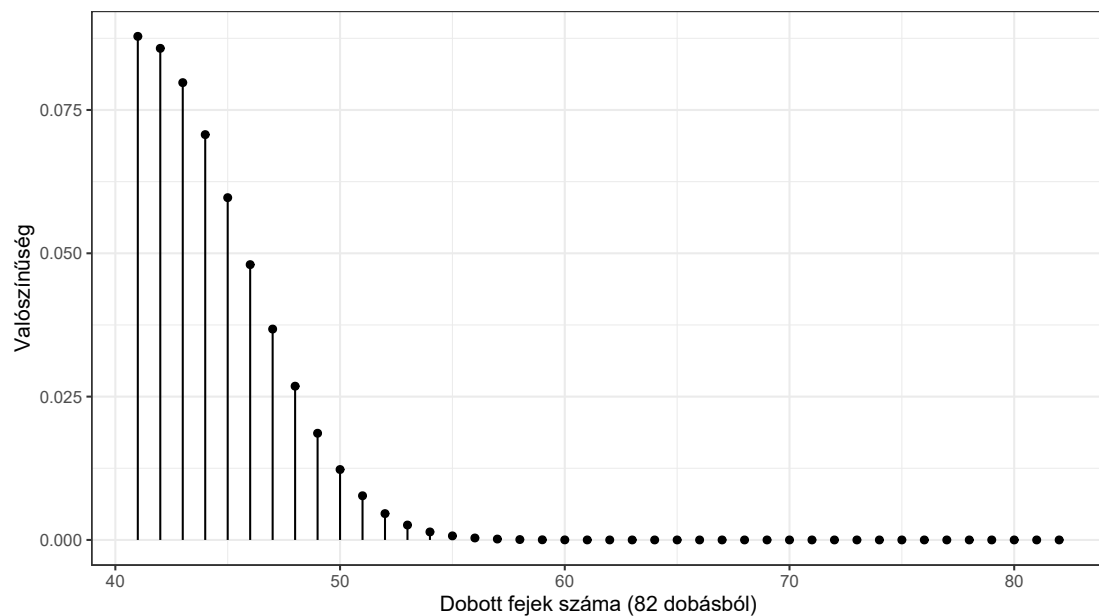
- Alapvető észrevétel: ha ragaszkodunk hozzá, hogy *biztos* döntést hozunk, akkor nem tudunk válaszolni...
- ...és ez *mindig* így van! Ha 820-ból 820 fej lett volna, *biztosan* akkor sem tudunk dönteni
- Az *egyetlen* biztos válasz, hogy „nem tudom”, minden máshoz *muszáj* valamennyi hibát vállalni

### Arbuthnot gondolatmenete

- Hiba vállalása tehát: akkor is azt mondjuk, hogy szabálytalan a pénzérme, ha valójában szabályos
- E nélkül ugyanis soha nem tudunk semmit mondani: ha csak akkor mondjuk, hogy szabálytalan, ha tényleg biztosan az, akkor soha nem mondhatjuk, mert szabályosság mellett is kaphatunk, nem nulla valószínűséggel, csupa fejet, akárhányszor is dobjuk fel
- Ami igazán fontos: ezt a hiba-vállalást tudjuk számszerűsíteni
- Ha csak akkor mondjuk, hogy szabálytalan a pénzérme, ha 82-ből 82 fejet dobunk (de már 81-nél is azt mondjuk, hogy „á, ez lehetett a véletlen miatt!”) akkor a fent említett nagyon kis hibavalószínűséget vállaljuk
- Ha már 81-nél is azt mondjuk, hogy „nem hisszük, hogy ez a véletlen miatt volt” (mert ha szabályos lenne, akkor ez annyira valószínűtlen lenne), akkor kicsit nagyobb a hiba ( $1,7 \cdot 10^{-23}$ )
- Ez mind számszerűen meghatározható, hiszen a szabályosság esetén tudjuk, hogy 50% valószínűséggel dobunk fejet, ezért – ha még azt is hozzátesszük, hogy függetlenek a dobások – ebből minden kiszámolható

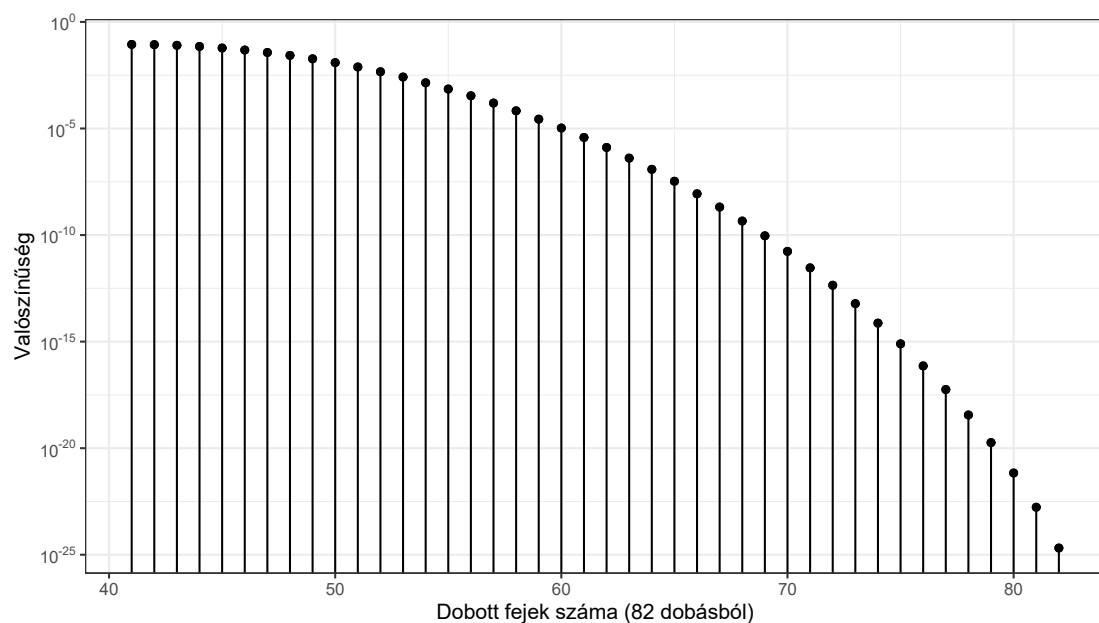
### Arbuthnot gondolatmenete

Számoljuk is ki ténylegesen: ha szabályos a pénzérme, ekkora valószínűséggel dobunk mégis, pusztán a véletlen ingadozás szeszélye révén adott számú fejet 82 dobásból



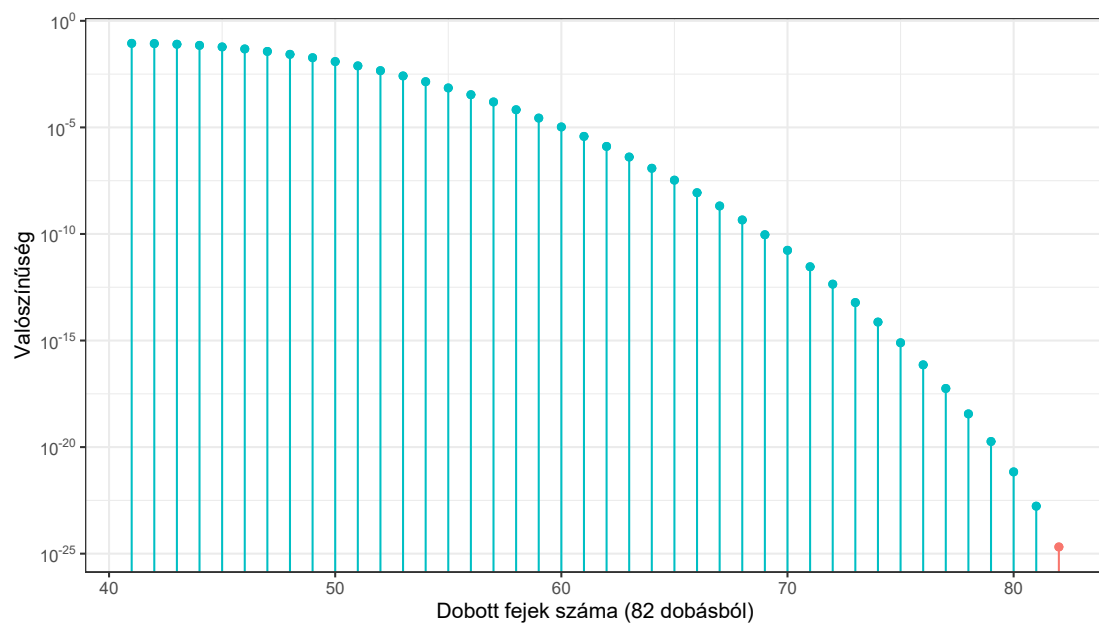
### Arbuthnot gondolatmenete

Számoljuk is ki ténylegesen: ha szabályos a pénzérme, ekkora valószínűséggel dobunk mégis, pusztán a véletlen ingadozás szeszélye révén adott számú fejet 82 dobásból



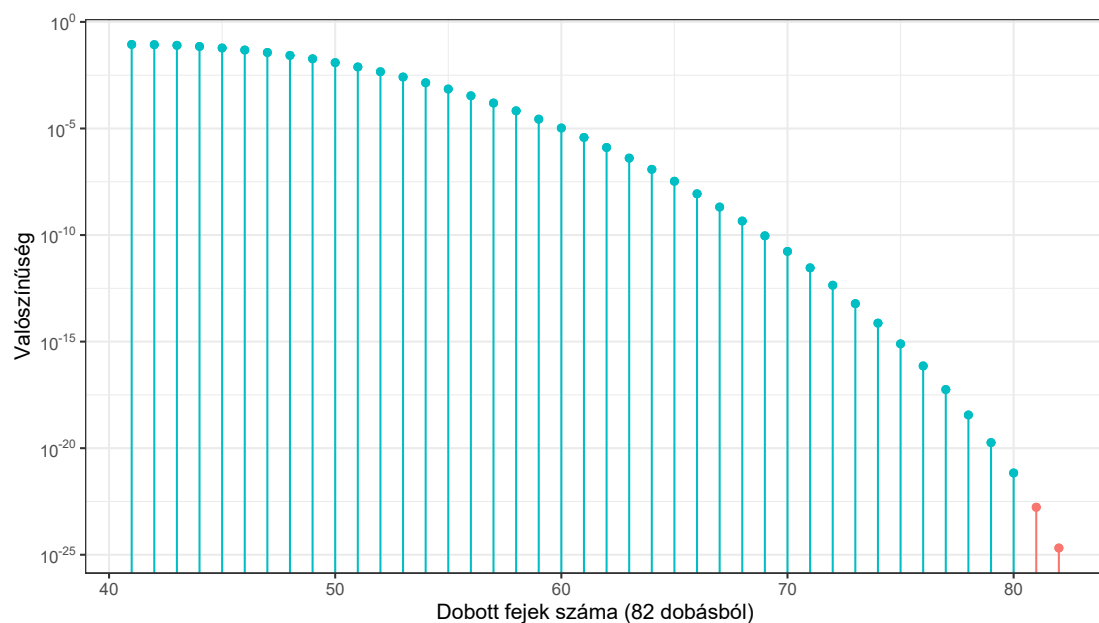
### Arbuthnot gondolatmenete

Ha csak 82-nél mondjuk, hogy szabálytalan:  $2,1 \cdot 10^{-25}$  – ilyen értelmű – hibavalószínűség



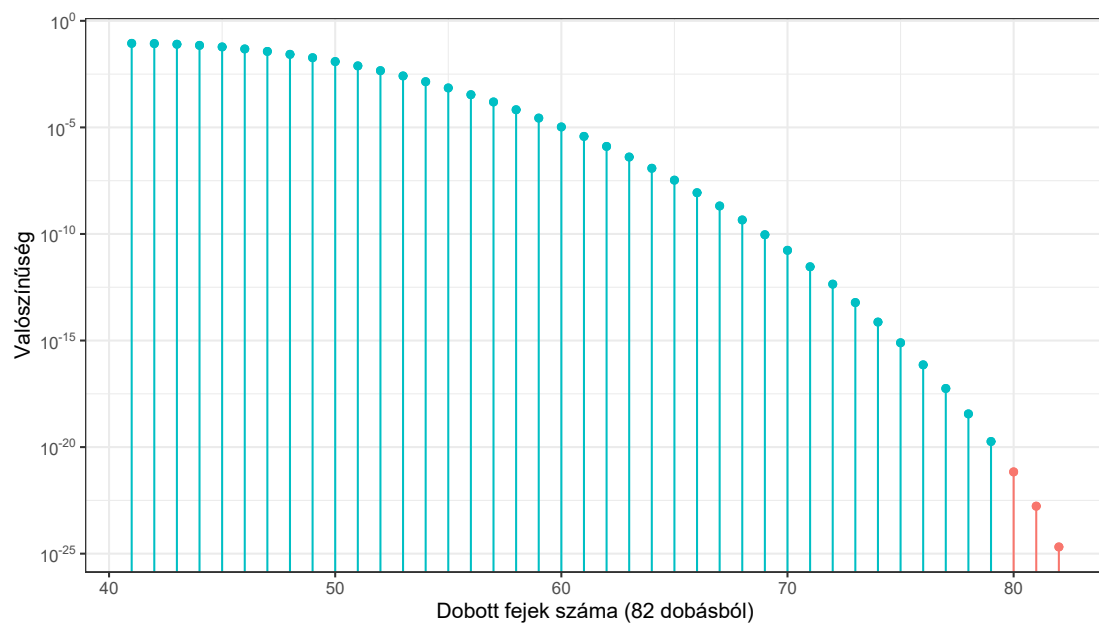
### Arbuthnot gondolatmenete

Ha 81-nél és 82-nél mondjuk, hogy szabálytalan:  $1,7 \cdot 10^{-23}$  – ilyen értelmű – hibavalószínűség



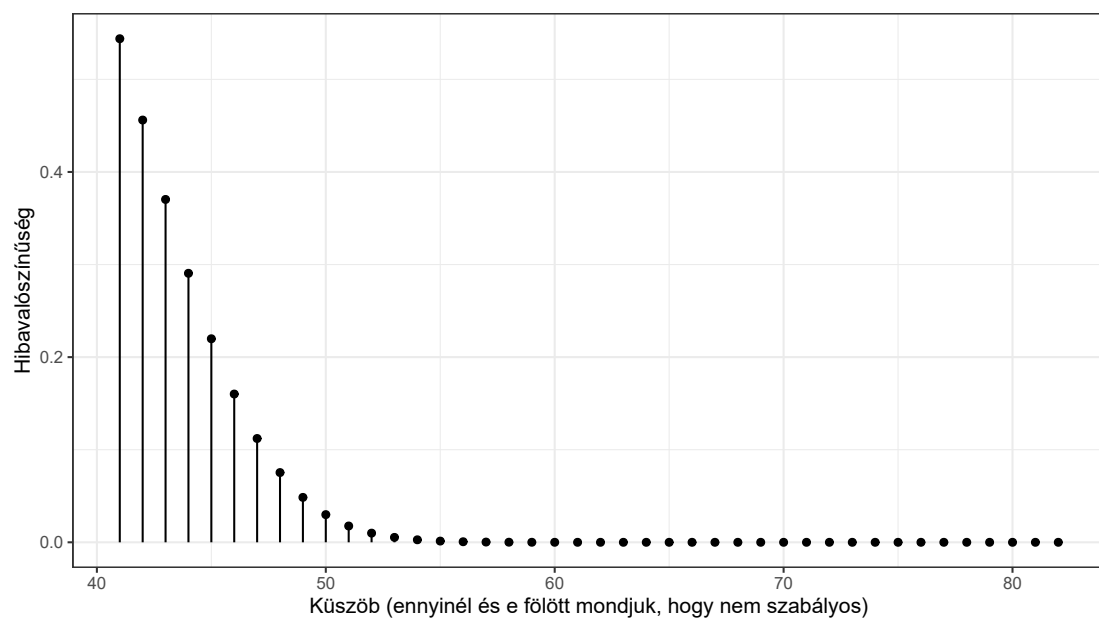
### Arbuthnot gondolatmenete

Ha 80-nál, 81-nél és 82-nél mondjuk, hogy szabálytalan:  $7,0 \cdot 10^{-21}$  – ilyen értelmű – hibavalószínűség



### Arbuthnot gondolatmenete

Mekkora hibát vállalunk (a „tévesen mondjuk, hogy szabálytalan a pénzérme (van Isten) miközben igazából nem” értelemben), ha adott küszöböt választunk



### Arbuthnot gondolatmenete

- Például úgy szeretnénk dönteni, hogy 5% hibát vállalunk ebben az értelemben



(tehát ha szabályos a pénzérme, 5% valószínűséggel mondjuk mégis, tévesen, a véletlen ingadozás szeszélye folytán, hogy nem szabályos)

- Kiszámolható, hogy ehhez mi lesz a megfelelő határ, lásd előző ábra
- A válasz 49 (kb. 5% hibavalószínűség)
- („kb.”: a diszkréttség miatt nem lehet itt most teljesen pontosan beállítani)
- Ha abban az esetben mondjuk, hogy nem szabályos a pénzérme, ha 49-szer vagy annál többször dobunk fejet 82-ből, akkor e stratégia mellett igazolhatóan 5% lesz a valószínűsége, hogy egy szabályosra azt mondjuk, hogy szabálytalan
- (Más szóval: ha nagyon sokszor kell ilyen helyzetben döntenünk, akkor a fenti szabályt követve hosszú távon az esetek 5%-ában fogunk ilyen értelmű hibát véteni)
- Ha nem 82-ször dobjuk fel a pénzérmét, akkor természetesen más lesz a döntési szabály, de a fenti módon meghatározható minden esetben a szükséges küszöb

### **Mi köze ennek az orvosi vizsgálatok kiértékeléséhez?**

Arbuthnot fenti modellje szinte tökéletesen megfelel a mai orvosi vizsgálatok kiértékeléséhez túlnyomó részében alkalmazott modellnek, csak a szavakat kell lecserélni:

- Pénzérme feldobása: adott beteg meggyógyul-e – fej: igen, írás: nem
- Pénzérme fej-dobási valószínűsége: gyógyulási valószínűség a gyógyszeres kezeléssel
- Tegyük fel, hogy gyógyszer nélkül a betegek fele meghal, és hogy a gyógyszer csak javítani tud (vagy nem változtat, ha hatástalan)
- (Valójában a gyógyszer nélküli arányt sem tudjuk köbevésetten – ezért kell két csoport – és az sem biztos, hogy csak javítani tud, de most egyszerűsítsünk, hogy a lényegre koncentráljunk)
- 82 betegem van, akinek gyógyszert adok, 82 meggyógyul – hat-e a gyógyszer?
- Tökéletes a megfeleltetés: hatástalan a gyógyszer, azaz 50% a kezelés melletti gyógyulási arány – szabályos a pénzérme, hatásos a gyógyszer, azaz 50%-nál nagyobb a gyógyulási arány – szabálytalan a pénzérme (figyelem, *bármennyivel* nagyobb)
- Mintanagyság – érmedobások száma
- Ugyanúgy csak hibával terhelt választ lehet hozni, *de* úgy, hogy a hiba mértéke maga is kontrollálható

### Mi köze ennek az orvosi vizsgálatok kiértékeléséhez?

- Ilyen értelmű hiba: a hatástalan gyógyszert tévesen hatásosnak hisszük
- Adja magát egy kérdés: ha ez hiba (ráadásul elég drámai, egy hatástalan gyógyszer törzskönyvezünk), és a mértékét mi szabjuk meg, akkor miért 5%-ot választottunk?? Miért nem 4-et? 1-et? Egymilliomodot??
- Nem beugratós a kérdés: ha ezt az értéket csökkentjük akkor *tényleg* csökken annak a valószínűsége, hogy hatástalan gyógyszert hatásosnak hiszünk...
- ...csak épp ezzel a hibázással szemben áll egy másik: hogy egy hatásost is hatástalannak hiszünk!
- Ha csökkentjük ezt a valószínűséget (nem elég 50 betegnek meggyógyulnia, 55-nek kell, 60-nak, stb.), akkor csakugyan ritkábban fogunk hatástalan gyógyszereket törzskönyvezni, de egyre gyakrabban fogjuk még a hatásosakra is nagy valószínűséggel azt mondani, hogy „á, ez lehetett a véletlen miatt” – erre a gondolatra mindjárt visszatérünk
- Vegyük észre, hogy ebben a gondolkodási keretben egy fordított logika érvényesül: nem azt nézzük, hogy ezen eredmény alapján mennyire valószínű, hogy hatástalan a gyógyszer, hanem azt, hogy ha hatástalan lenne, akkor mennyire valószínű, hogy ilyen eredményt kapnánk

### Precíz terminológia

- Nullhipotézis (null hypothesis)
  - Kiinduló, „alapértelmezett” állításunk, amit további információ nélkül elhiszünk
  - Tipikusan azt fogalmazza meg, hogy nem találtunk semmi újat (nincs eltérés, nincs hatás, nincs különbség)
  - Szabályos a pénzérme, nincs különbség a fiú-lány arányban, nem hat a gyógyszer
  - Lényegében van egy kiindulópontunk és azt nézzük, hogy a begyűjtött adat mennyire mond neki ellent
- Ellenhipotézis (vagy alternatív hipotézis, alternative hypothesis)
  - A nullhipotézis tagadása
  - Újdonság, heuréka, találtunk valamit (van eltérés, van hatás, van különbség)
  - Nem szabályos a pénzérme, eltér a fiúk és a lányok aránya, hat a gyógyszer
- A nullhipotézis általában csak egyféleképp tud teljesülni, az ellenhipotézis több módon is (a szabályosság 50% fejdobási valószínűség és csak az, szabálytalan sokféleképp lehet; a hatástalanság csak a 0 hatás, hatásos sokféleképp lehet, hathat kicsit is, meg nagyon is stb.)

## Precíz terminológia

- Szignifikanciaszint (significance level): a célvalószínűség, ami alapján meghatározzuk a küszöb helyét; jele:  $\alpha$
- Elsőfajú hiba (Type I error)
  - Tévesen minősítünk hatásosnak egy hatástalan gyógyszert
  - Elvetjük a nullhipotézist, pedig igaz
  - Ha teljesülnek a próba előfeltevései, akkor a valószínűsége épp az általunk választott szignifikanciaszint,  $\alpha$
  - Azaz: a valószínűsége felett erős kontrollunk van (semmi egyébtől nem függ – természetesen a döntési szabály függhet például a mintanagyságtól, de az elsőfajú hiba valószínűsége nem)
- Másodfajú hiba (Type II error)
  - Tévesen minősítünk hatástalannak egy hatásos gyógyszert
  - Nem vetjük el a nullhipotézist, pedig nem igaz
  - A valószínűségének a jele:  $\beta$  (annak a valószínűsége, hogy nem vetjük el a nullhipotézist, feltéve, hogy nem igaz)
  - $1 - \beta$  az erő: feltéve, hogy hat a gyógyszer, mekkora valószínűséggel mutatjuk is ezt ki
  - Nincs felette erős kontrollunk

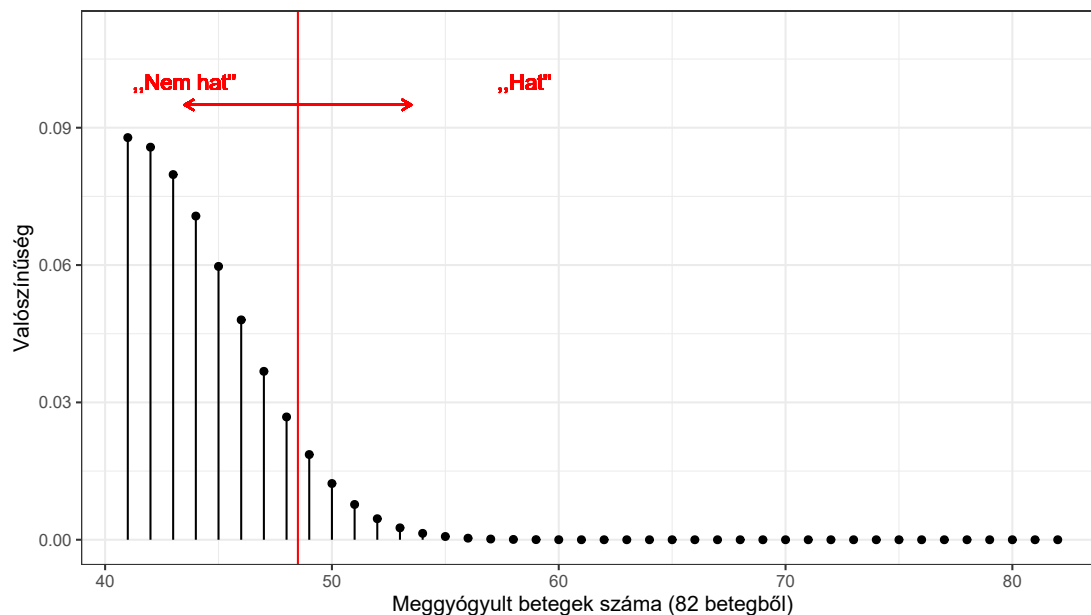
## Precíz terminológia

- Szignifikáns hatás: adott szignifikanciaszinten elvetjük a nullhipotézist
  - Tartalmilag: a tapasztalt hatás – 42-nél többen gyógyultak meg – betudható-e egyszerűen a véletlen ingadozásnak, tehát a valóságban nem hat a gyógyszer és ami hatást láttunk az csak ebből a véletlenségből fakadt, vagy olyan nagy, hogy azt feltételezzük, hogy valódi különbség van mögötte (tényleg hat a gyógyszer)
- A kérdésünk az lesz, hogy a begyűjtött adatok „elégé” ellentmondanak-e a nullhipotézisnek (azaz: ha a nullhipotézis fennállna, akkor valószínűtlen, hogy az történjen, ami történt), ahhoz, hogy azt mondjuk, hogy „nem hisszük el” (nem kizárjuk!), hogy fennáll (vegyük észre itt is a fordított logikát!)
- Vagy azt mondjuk, hogy ez csak olyan kevésbé tér el, hogy ezt még „hajlandóak vagyunk” betudni a véletlen ingadozásnak
- (A „mit hiszünk el”, „mit vagyunk hajlandóak betudni” dolgot szabályozza a szignifikanciaszint)

- „Elfogadjuk”: elfogadjuk a nullhipotézist (azaz azt mondjuk, hogy nem találtunk hatást, pl. nem hat a gyógyszer) – bár jobb lenne úgy fogalmazni, hogy „nem tudunk elutasítani”, később majd kiderül, hogy miért
- „Elutasítunk”: elutasítjuk a nullhipotézist (azaz azt mondjuk, hogy találtunk hatást, pl. hat a gyógyszer)

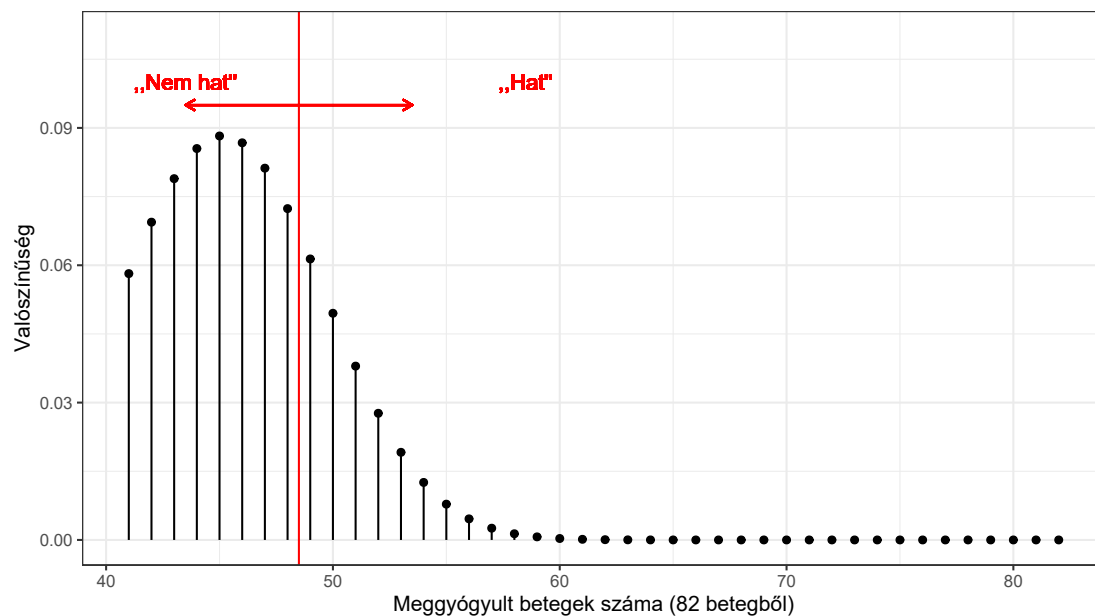
### Az erő és a hatásosság összefüggése

Mi az, hogy erő? Vegyük elő az ábrát hatástalanság esetén (ez ugye még nem jó az erőhöz!):



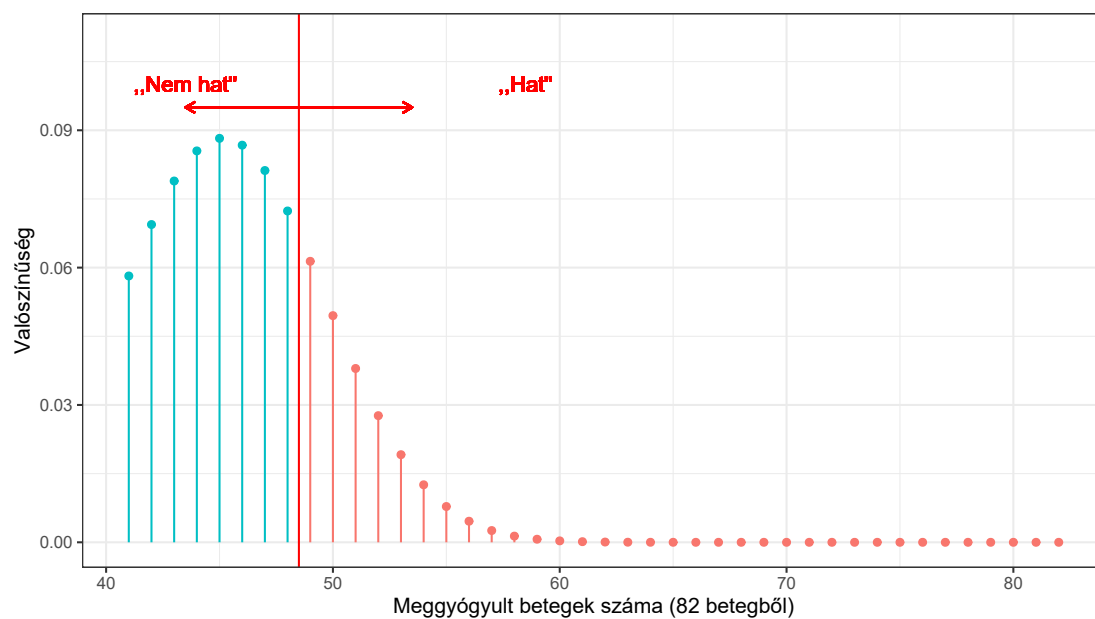
### Az erő és a hatásosság összefüggése

Mi az, hogy erő? Feltéve, hogy hat a gyógyszer (50-ről 55%-ra emeli a gyógyulási arányt)



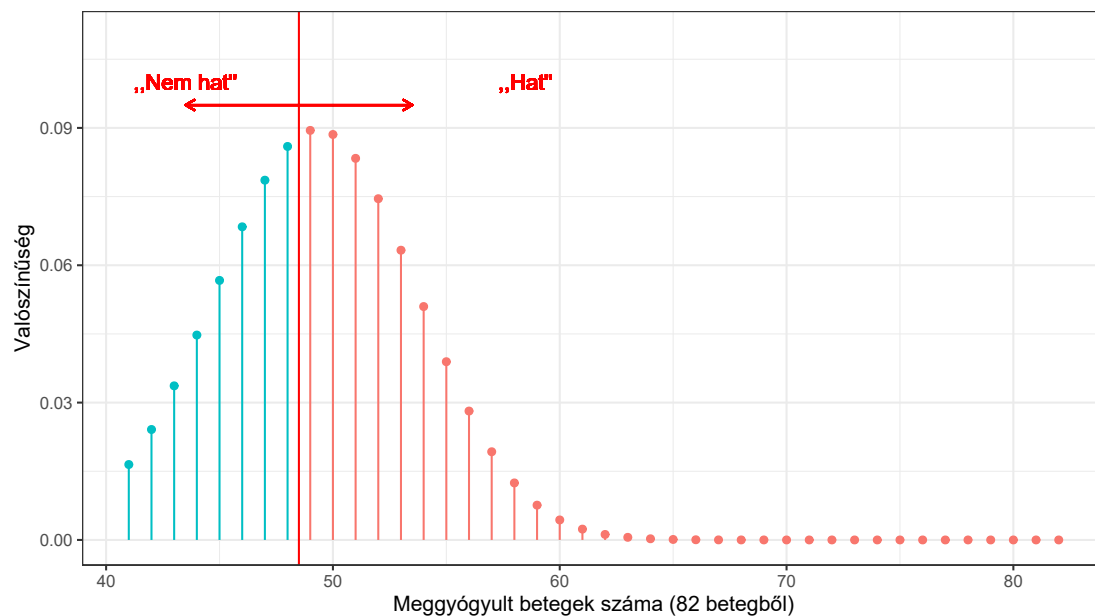
### Az erő és a hatásosság összefüggése

Mi az, hogy erő? Feltéve, hogy hat a gyógyszer (50-ról 55%-ra emeli a gyógyulási arányt), mekkora valószínűséggel mondjuk mi is azt, hogy hat



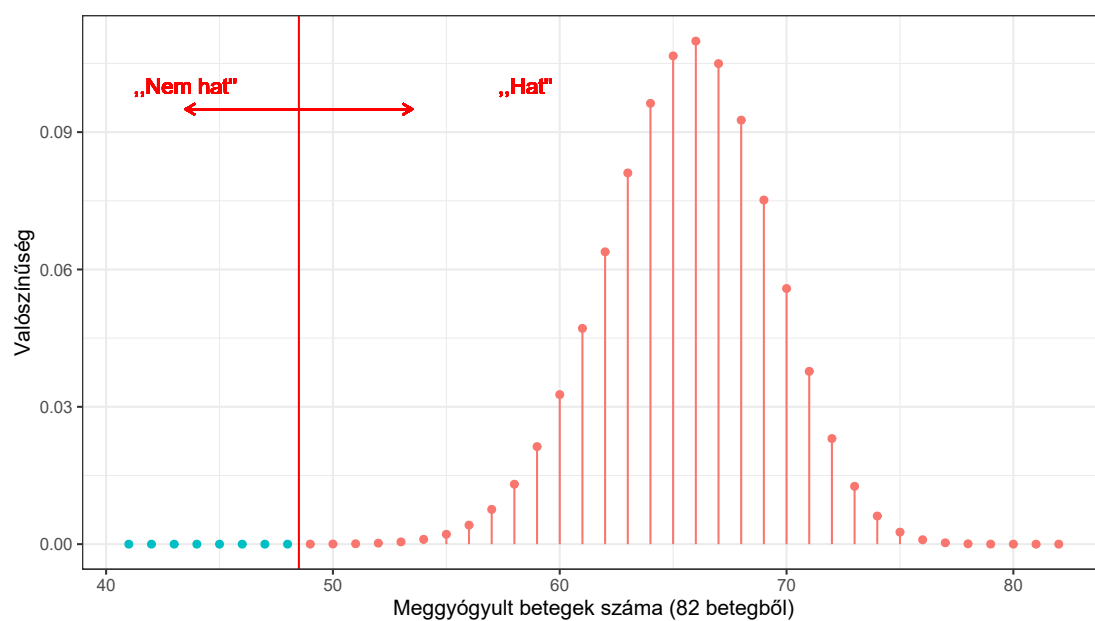
### Az erő és a hatásosság összefüggése

Kicsit jobban ható gyógyszer (60%)...

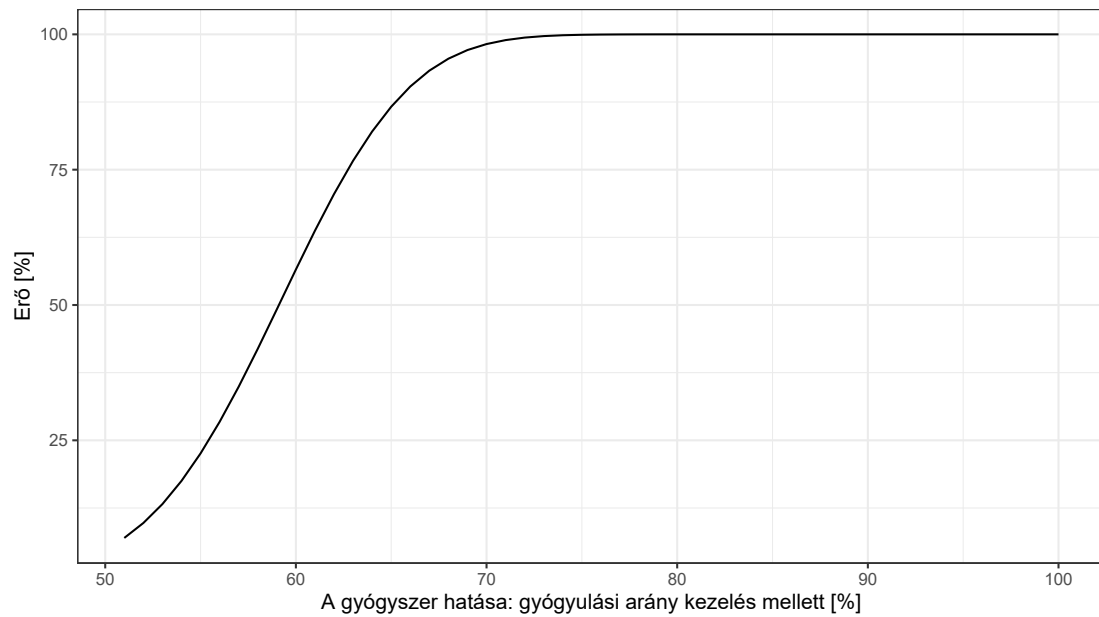


### Az erő és a hatásosság összefüggése

Még jobban ható gyógyszer (80%)...



### Az erő és a hatásosság összefüggése



### Az erő és a hatásosság összefüggése

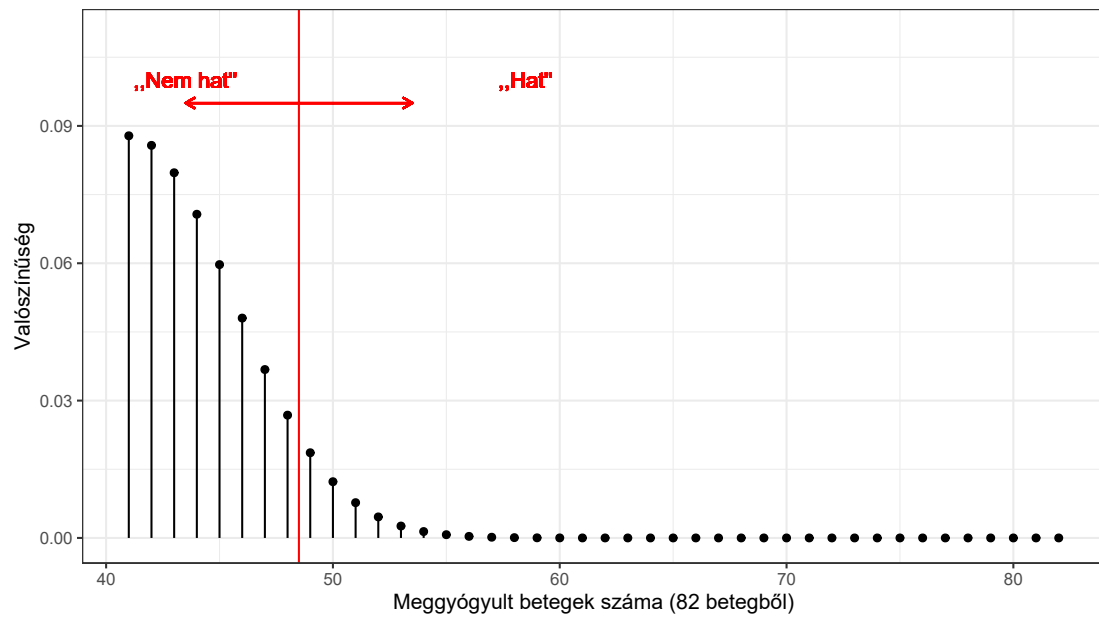
- Természetesen a döntési határ minden esetben ugyanott van (hiszen azt a hatástalanság feltételezése melletti eloszlás alapján határoztuk meg, nem függ a gyógyszer hatásától!)
- Tanulság: az erő függ a gyógyszer valódi hatásától – minél hatásosabb a gyógyszer, annál nagyobb lesz (minden mást változatlanul tartva) az erő

### Az erő és a mintanagyság összefüggése

- Hogyan tudjuk növelni az erőt? (hiszen a gyógyszer hatásosságára nincs ráhatásunk)
- Intuitív ötlet: növeljük meg a mintanagyságot! („kiátlagolódnak az eltérések”)
- Csakugyan így van, nézzük meg miért

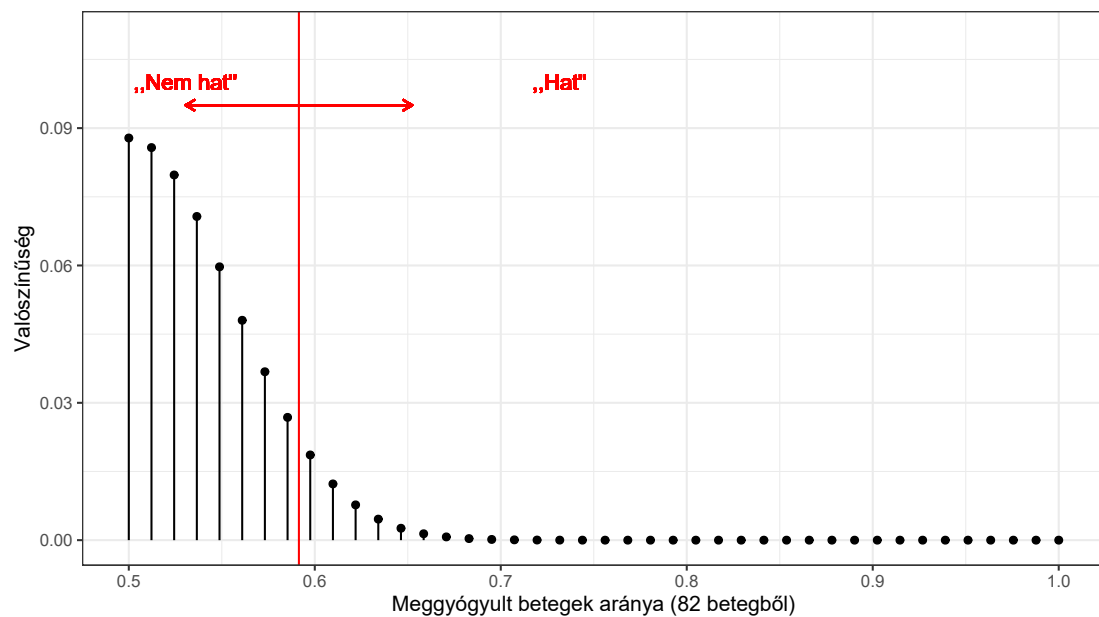
### Az erő és a mintanagyság összefüggése

Az eredeti ábránk (82 beteggel)



### Az erő és a mintanagyság összefüggése

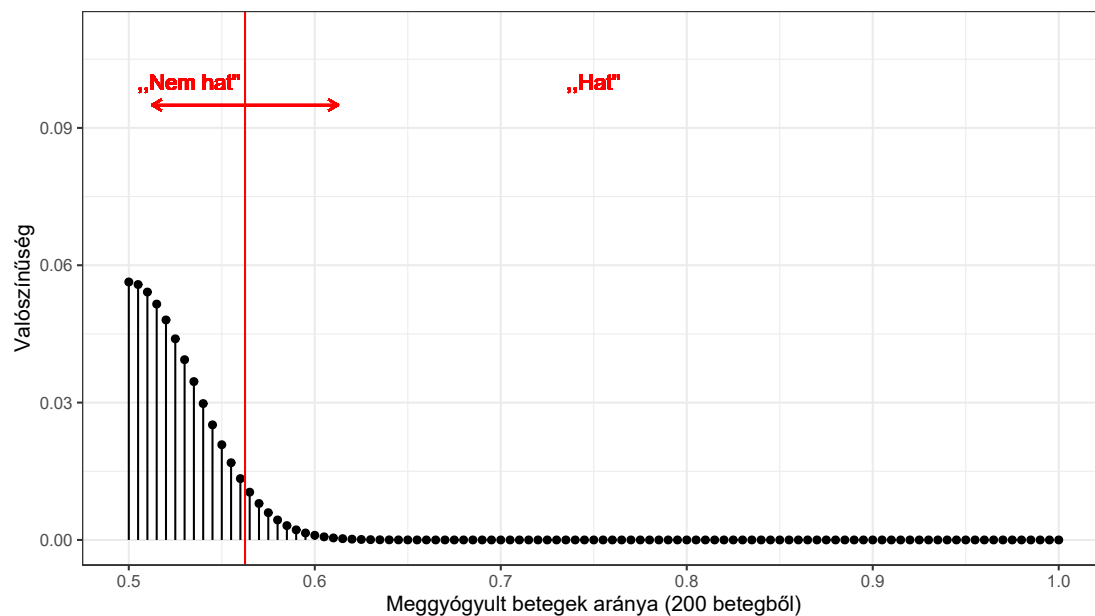
A vízszintes tengelyt skálázzuk át (csak hogy összehasonlítható legyen később)



### Az erő és a mintanagyság összefüggése

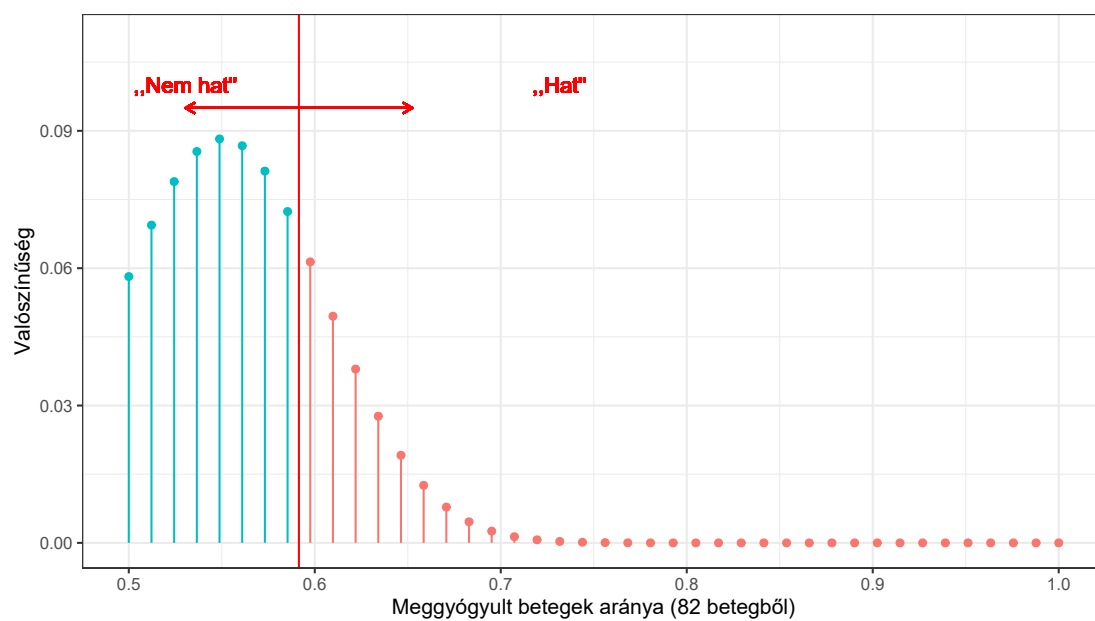
Ugyanez az ábra 200 beteggel (mi változott?)





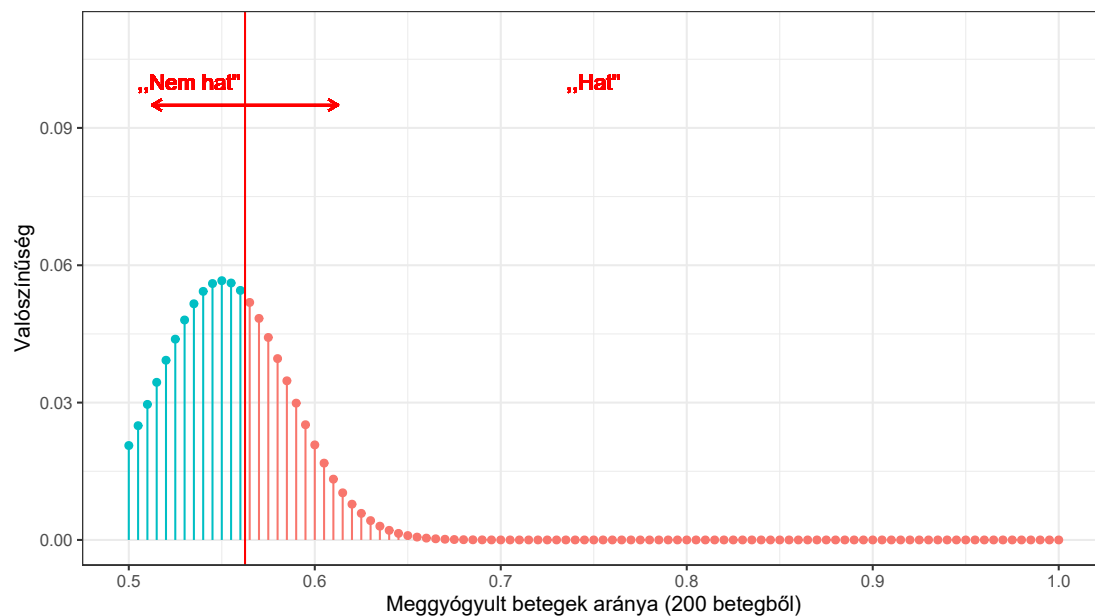
### Az erő és a mintanagyság összefüggése

Kicsit ható gyógyszer, 82 beteg



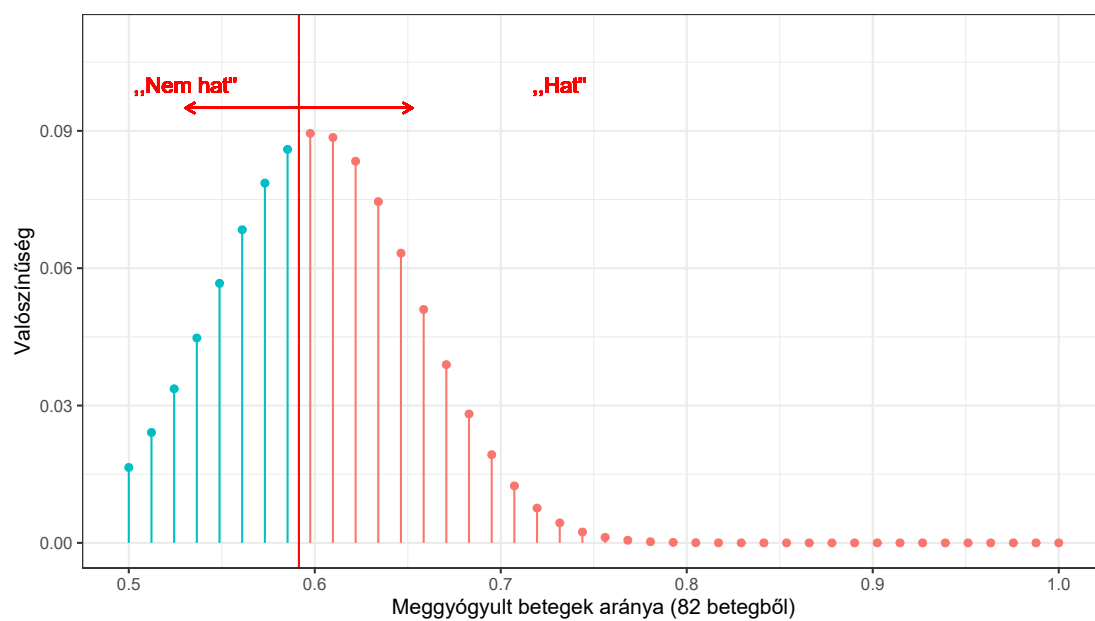
### Az erő és a mintanagyság összefüggése

Kicsit ható gyógyszer, 200 beteg



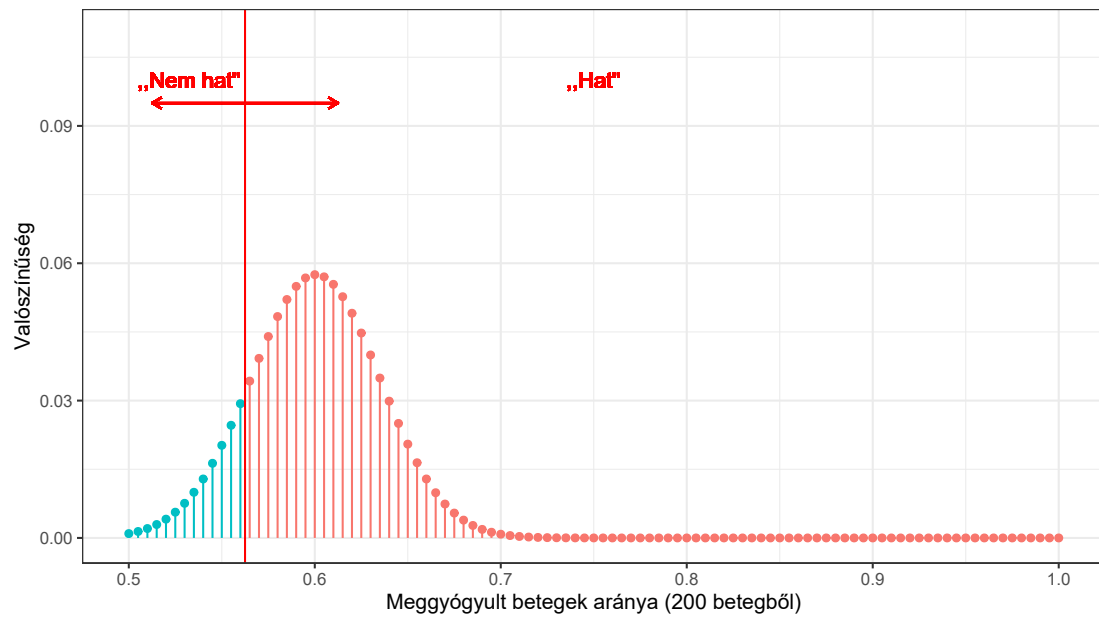
### Az erő és a mintanagyság összefüggése

Közepesen ható gyógyszer, 82 beteg



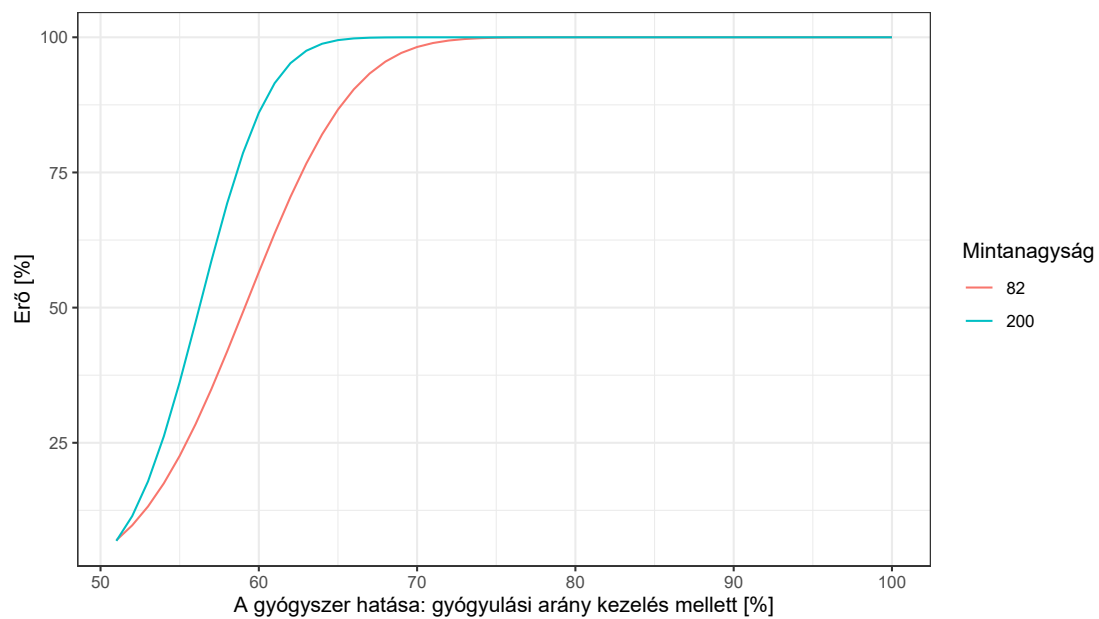
### Az erő és a mintanagyság összefüggése

Közepesen ható gyógyszer, 200 beteg



### Az erő és a mintanagyság összefüggése

A végeredmény:

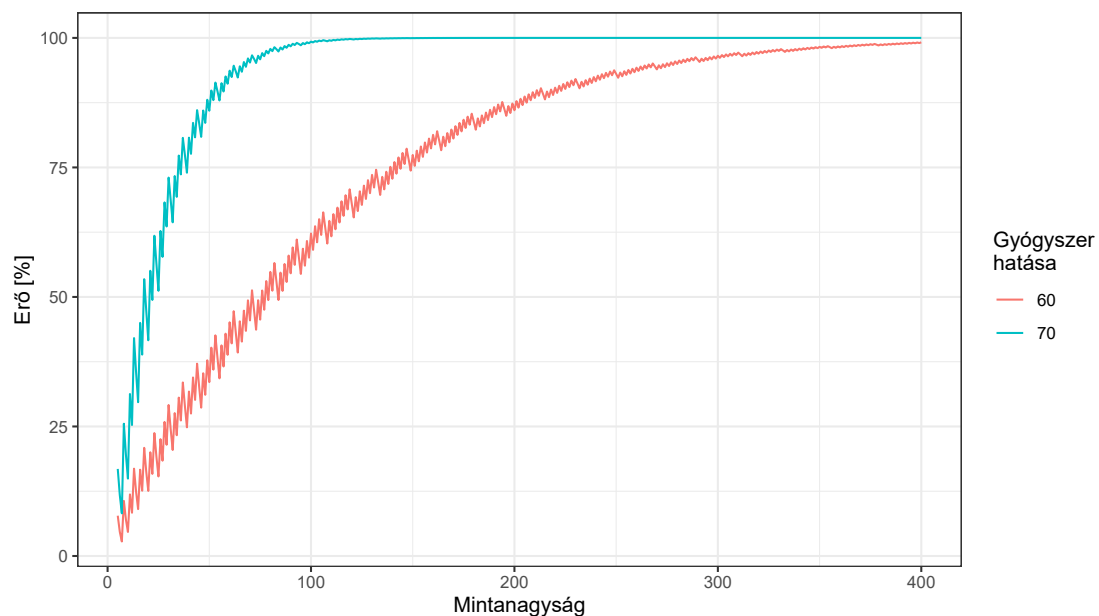


### Az erőt meghatározó tényezők

- Végeredményben tehát azt látjuk, hogy igazából két tényező hat az erőre: a gyógyszer hatásossága (hatásnagyság) és a mintanagyság

- (Sőt, valójában három, mert a szignifikanciaszint is)
- Kétféleképp is elmondható: ugyanazon hatásosság mellett nagyobb az erő, ha nagyobb a mintanagyság, *avagy* kisebb hatás is kimutatható ugyanazon erővel, ha nagyobb a mintanagyság
- „Kis hatásához nagy minta kell, nagy hatásához kis minta is elég”
- (Mármint a kimutatásához, értsd: hogy igazolni is tudjuk, hogy az nem csak a véletlen ingadozás miatt van, tehát, hogy szignifikáns)
- Nézzük meg úgy is ezt a függvényt, hogy a hatást rögzítjük, és a mintanagyságot variáljuk

### Az erő és a mintanagyság összefüggése



### Racionális mintanagyság-tervezés, erőelemzés

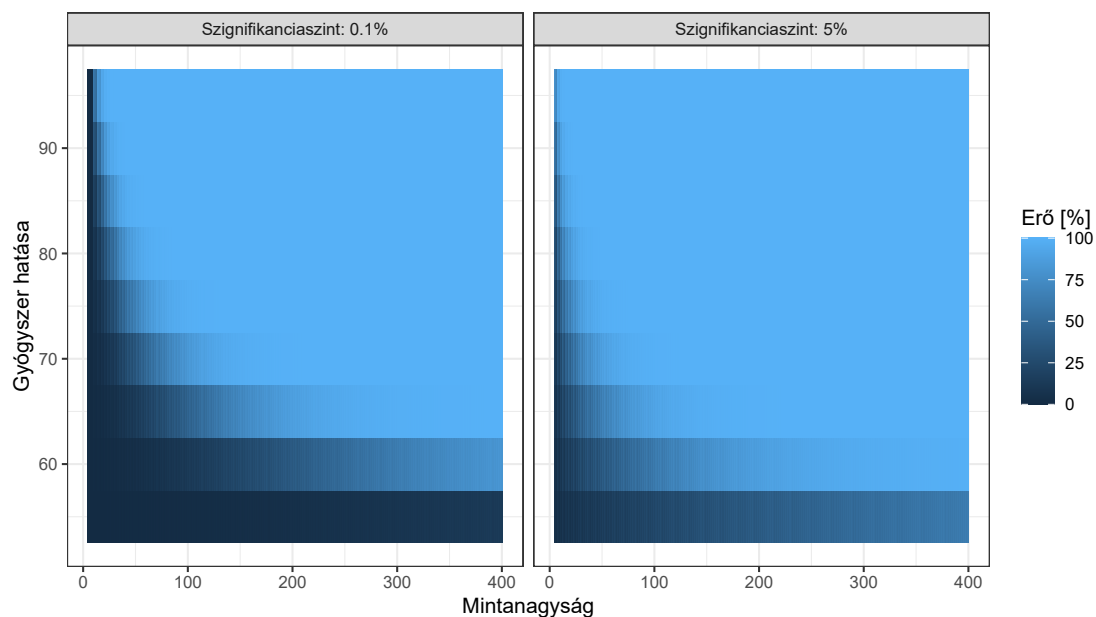
- (A furcsa szaggatás a diszkréttség miatt van – nem tud 70,2 beteg meggyógyulni – ne törődjünk vele)
- Ez egy nagyon fontos dolgot lehetővé tesz: a mintanagyság racionális tervezését!
- Csak az előbbi ábrára „fordítva” nézünk rá: nem mintanagysághoz olvassuk le az erőt, hanem adott, megkívánt erőhöz – tipikusan 80% – keressük ki a mintanagyságot

- (Ellenőrző kérdés: miért 80? Miért nem 90? 99? – hiszen akkor jobb eséllyel igazolni is tudjuk, hogy hat a gyógyszer, ha tényleg hat...!)
- Figyelem: minden ilyen számítás tehát feltételes azon, hogy mekkora a gyógyszer hatása!
- Furcsa lehet, hogy előre meg kell adnom a hatást, hiszen a kutatás célja épp ennek kiszámolása, de ez olyan mint a csillagászat: minél halványabb csillagot akarok kimutatni, annál nagyobb távcső kell – úgyhogy amikor távcsövet veszünk, kell valamilyen feltételezéssel élni a csillag fényességéről

### Racionális mintanagyság-tervezés, erőelemzés

- Kétféleképp is elmondható: ugyanazon mintanagyság mellett nagyobb az erő, ha hatásosabb a gyógyszer, *avagy* ugyanazon erőhöz kisebb minta is elég, ha hatásosabb a gyógyszer
- (Látszott az előző ábrán is: igazából ugyanazt a háromdimenziós függvényt mondjuk el, csak kérdés, hogy hogyan nézünk rá, milyen metszetet készítünk belőle)

### Az erőt meghatározó tényezők



### Az erőt meghatározó tényezők

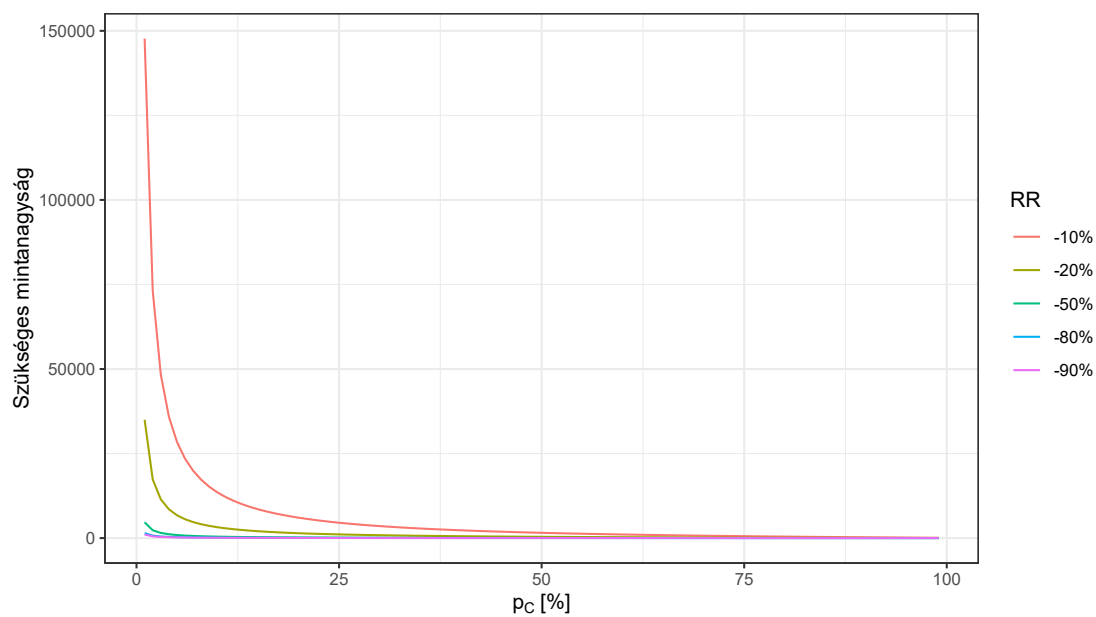
- Itt most egyszerre látjuk az összes tényezőtől való függést
- Bejött a szignifikanciaszint is

- Tanulság: ha lecsökkentjük, leromlik az erő is
- (Emlékezzünk vissza mit mondtunk már a legelején: ha nagyon szigorúak vagyunk, akkor a valódi hatást is nehéz lesz észrevenni...)

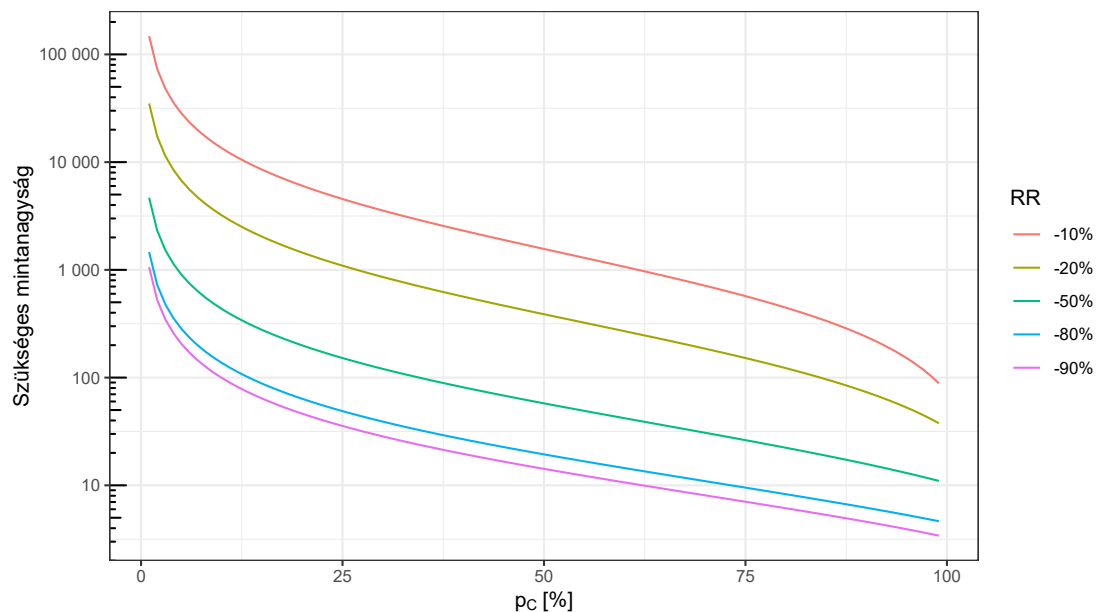
### Egy valódi esettanulmány

- Igazi példa: most már két csoport van, gyógyszer nem csak javítani tud
- Profilaktikus példa: a kontrollcsoportban sem biztos, hogy mindenki elszenvedi a végpontot
- A kontrollcsoportban  $p_C$  a kockázat, a kezelés ezt 10, 20, 50, 80 illetve 90%-kal csökkenti
- Mekkora mintanagyság kell ennek a kimutatásához?
- (Szokásosan  $\alpha = 0,05$  szignifikanciaszint,  $1 - \beta = 0,8$  erő)

### Egy valódi esettanulmány



### Egy valódi esettanulmány



### Egy valódi esettanulmány

- Gondoljuk végig az eddigieket az ábrán!
- Egy plusz tanulság: annál könnyebb szignifikánsnak kimutatni valamit, minél nagyobb az abszolút különbség is a csoportok között (még ha a relatív különbség állandó is!)
- Lásd vakcinás példa:
  - Ezért hatalmasak az ilyen kísérletek
  - Persze kölcsönhat az utánkövetés hosszával (abszolút húsbavágó kérdés: lehet, hogy kevesebb alannyal is jók lennénk, ha hosszabb a vizsgálat, *vagy* ugyanannyi alannyal pontosabb eredményt kapnánk, *de* így később lenne csak eredmény...!)
  - Challenge kísérlet?

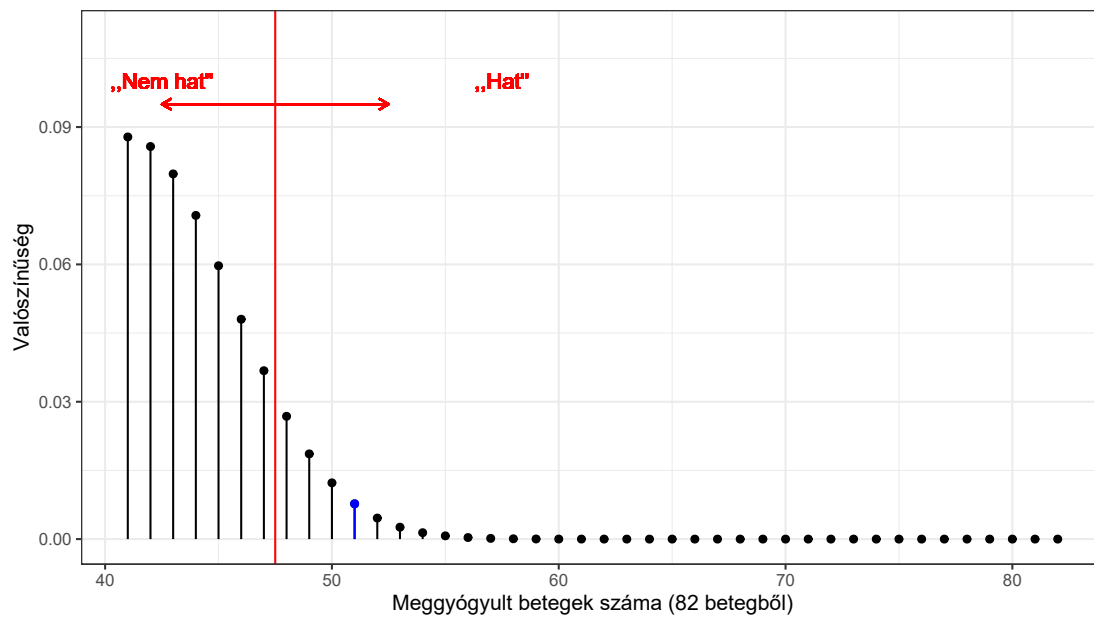
### A $p$ -érték

- A fentiek azt sugallják, hogy a szignifikanciaszinttel egy bináris döntést adjunk meg („hat” – „nem hat”)
- De így semmi nem derül ki abból, hogy mennyire erős az a döntés
- Vigyázat, az „erős” most azt jelenti: mennyire mond ellent a nullhipotézisnek (hatástalan gyógyszer), azaz mennyire vetjük el, hogy amit kaptunk, az pusztán a véletlen műve, miközben valódi hatás nincs

- Ötlet: ha nem csak azt adjuk meg, hogy 5%-on hat-e, hanem azt is, hogy 5%-on hat-e és 1%-on hat-e, akkor gazdagítjuk az információt
- És 0,1%-on? Vagy ha 5-ön sem hatott akkor 10-en...? És így tovább
- Például: 82-ből 51 gyógyult meg

#### A $p$ -érték

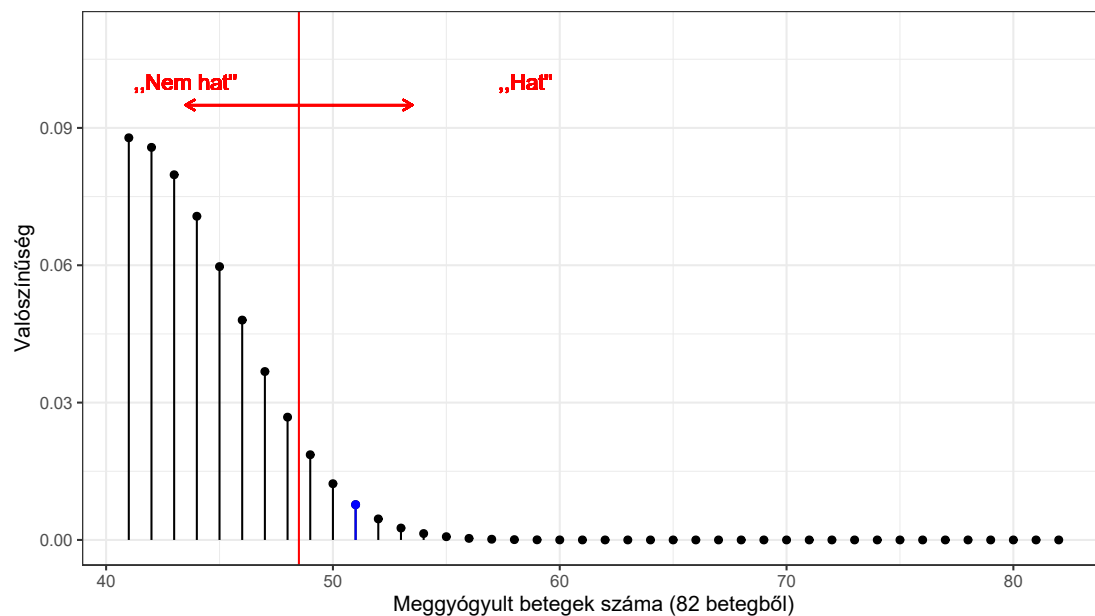
10%-os szignifikanciaszint: túl magas



#### A $p$ -érték

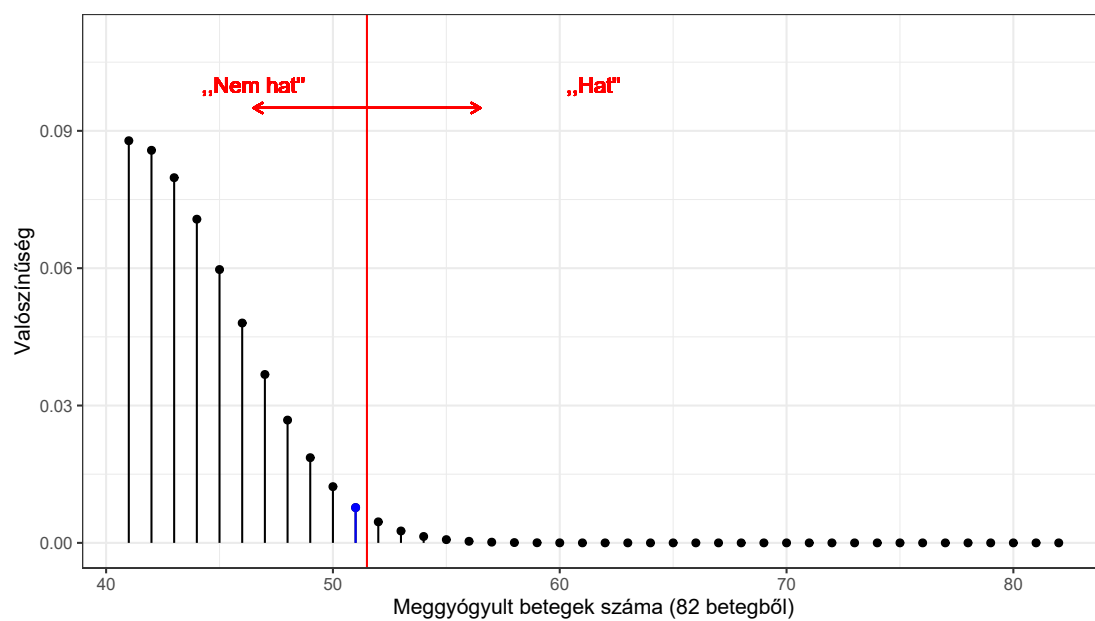
5%-os szignifikanciaszint: még mindig kicsit magas





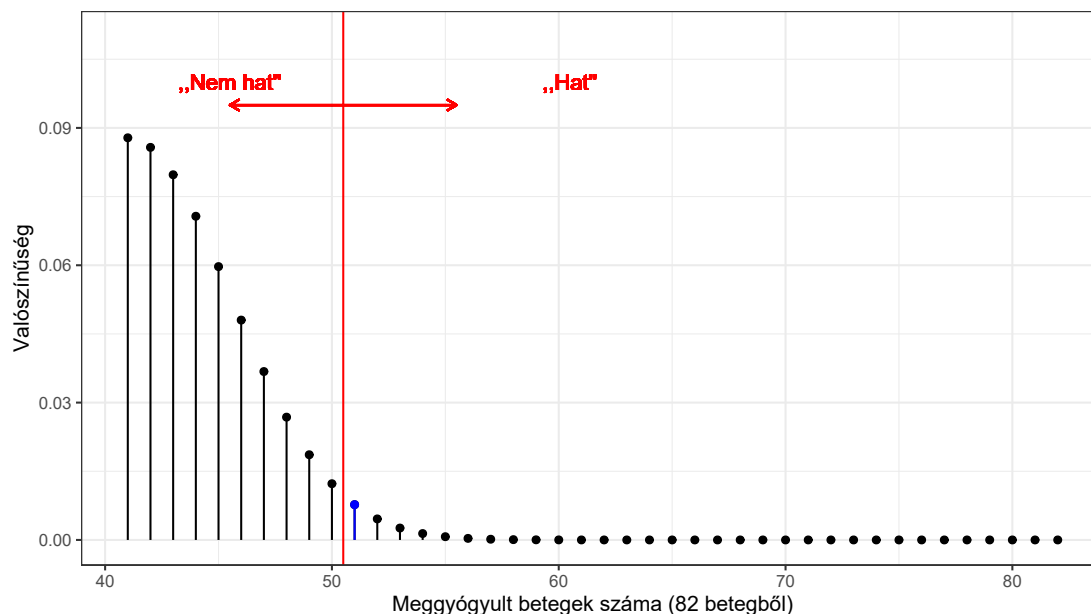
#### A $p$ -érték

1%-os szignifikanciaszint: hopp, ezzel már túllóttünk



#### A $p$ -érték

2%-os szignifikanciaszint: ez megint picit sok



### A $p$ -érték

- A pontos érték (egy lehetséges, népszerű módszerrel – a diszkréttség miatt nem egyértelmű)  $p = 0,01762$
- Ez tehát az a szignifikanciaszint, ahol épp az elfogadás és az elutasítás határán vagyunk
- Azaz: ha ennél nagyobb szignifikanciaszintet választunk, akkor elutasítunk (hatásosnak minősítjük a gyógyszert), hiszen ekkor „enyhítettünk”, ha kisebbet, akkor elfogadunk (hatástalannak minősítjük a gyógyszert), mert ekkor „szigorítottunk”
- (Ezért szokták mindig azt nézni, hogy a  $p$  kisebb-e mint 0,05, hiszen 5% a szignifikanciaszint szokásos értéke)
- Lényegében lehetővé teszi, hogy az olvasó tetszőleges, általa szimpatikusnak talált szignifikanciaszinten döntsön (ne csak a szerző által használaton, amit jelentene az, ha csak a döntés van leírva adott szignifikanciaszinten)

### A $p$ -érték másik interpretációja

- Egyúttal egyfajta mérőszáma is annak, hogy az adatok mennyire mondanak ellent a nullhipotézisnek (mennyire erősen bizonyítják, hogy a hatás nem pusztán a véletlen ingadozás miatti – természetesen a fordított logika értelmében!)
- Így már az is látszik, hogy az erősség alatt pontosan mit ért a  $p$ -érték: ha hatástalan lenne, akkor mennyire valószínű, hogy olyan, vagy annál is extrémebb

(hatástalanságnak még jobban ellentmondó) eredmény jöjjön ki, mint ami ténylegesen kijött

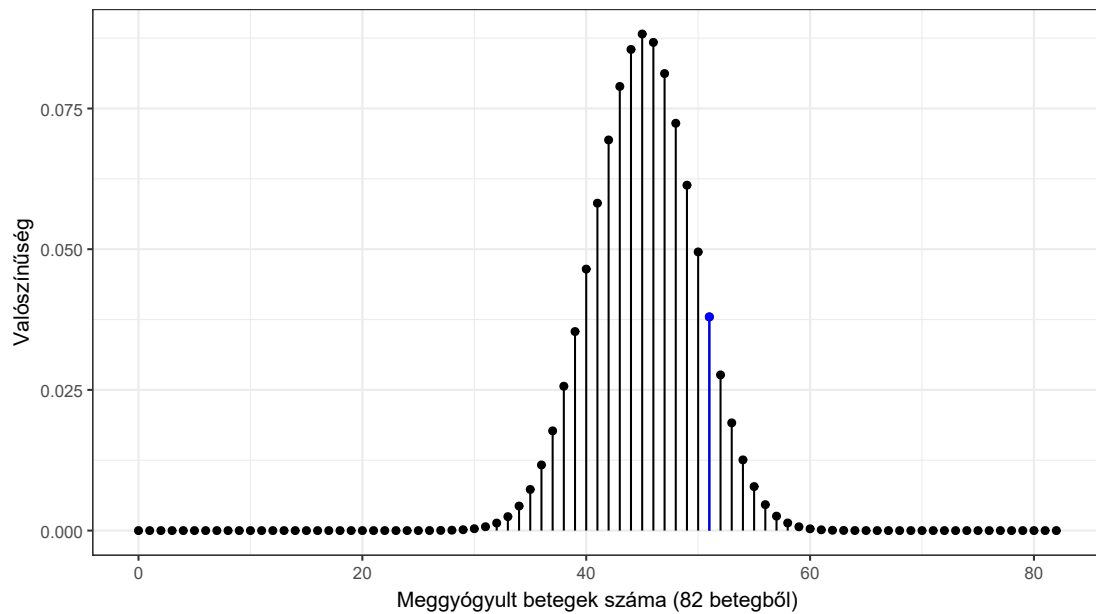
- Vigyázat, nem azt méri, hogy a gyógyszer mennyire hatásos, hanem azt, hogy mennyire biztosak vagyunk abban, hogy a tapasztalt hatás a véletlen műve – a kettő nem ugyanaz! (kis hatás is lehet nagyon biztosan a véletlen műve – mikor is? – és egy nagy hatás is lehet, hogy könnyen kijöhetett véletlenül)
- Sajnos itt két dolog keveredik, a manapság használt kiértékelési módszertan több iskola nem teljesen konzisztens keveréke

### Egy új kérdés

- A fentiekben végig azzal foglalkoztunk, hogy egy valóságra vonatkozó állítás („a gyógyszer valódi hatása nulla, mert a gyógyulási arány kezelés mellett is 50%”) igazságtartalmát vizsgáltuk
- Amint láttuk, „eldönteni” nem lehet, mindenképp bizonytalan lesz a válaszunk
- De feltehető ez a kérdés is: mit gondolunk, *mennyi* a gyógyszer hatása (pl. a kezelés melletti gyógyulási arány)...?
- Ahogy az előbbi kérdésre nincs biztos igen/nem válasz, úgy erre is csak bizonytalan, potenciálisan hibával terhelt válasz adható
- Például a 82 betegből 51 gyógyult meg – mit *tippelünk*, mennyi a gyógyulási arány?
- $51/82 = 62,2\%$ ?
- A legjobb *tippünk* valóban ez (ezt be lehetne matematikailag is bizonyítani, ha alkalmasan definiálnánk egy tippelés jóságát, most ezzel nem foglalkozunk)...
- ...de ez akkor is csak *tipp*!

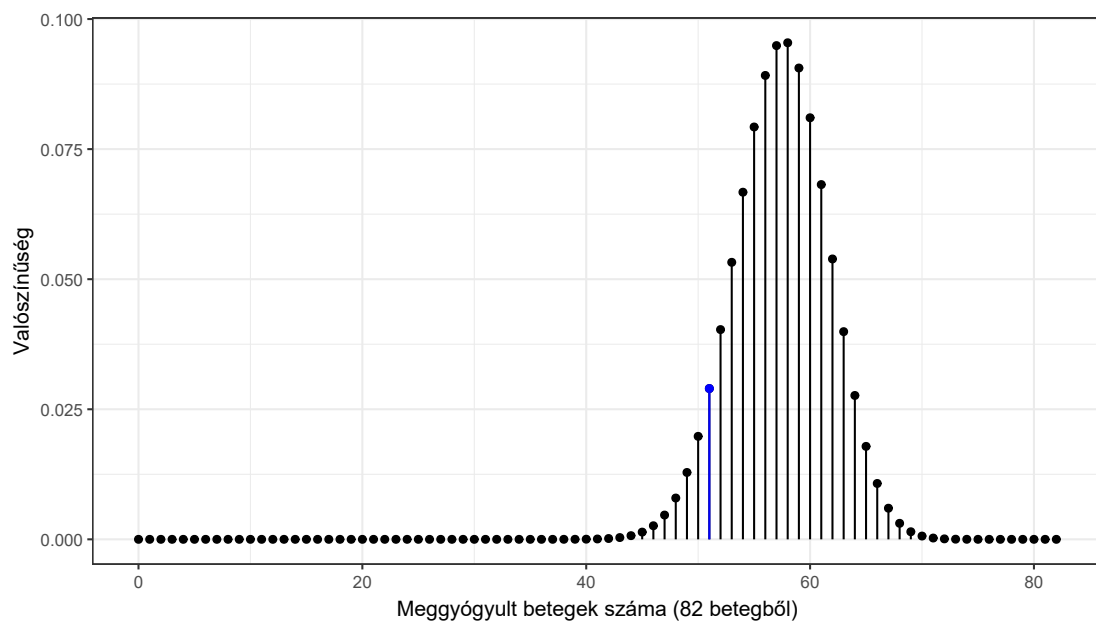
### Tippelés, mert...

Ha a valódi gyógyulási arány 55%, akkor kijöhet kényelmesen az 51/82, pusztán a véletlen ingadozás miatt...



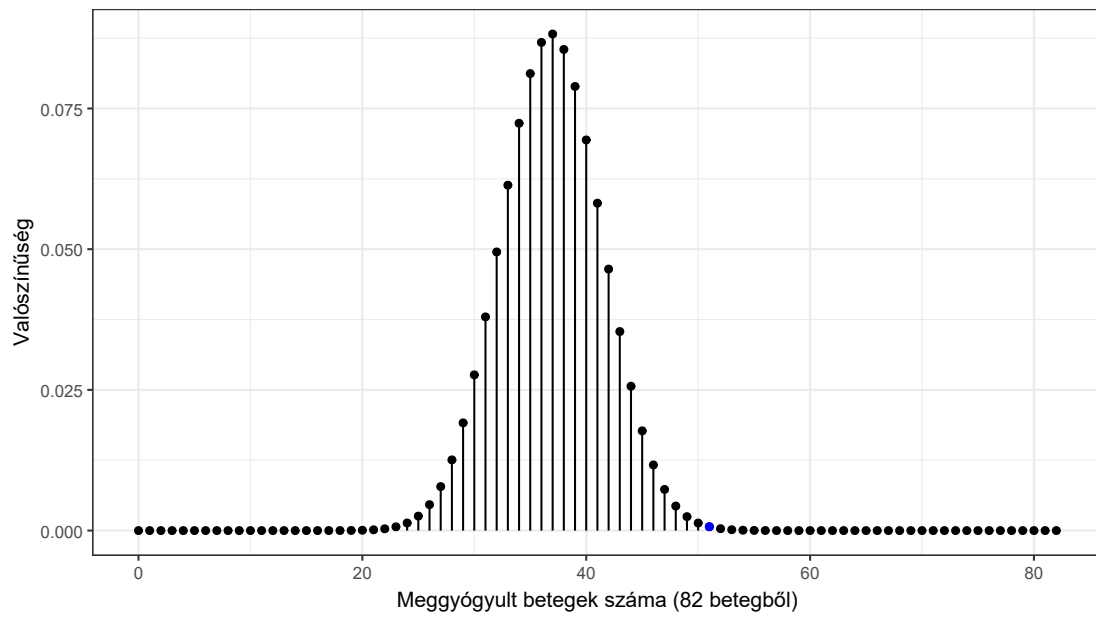
### Tippelés, mert...

Ha a valódi gyógyulási arány 70%, akkor is kijöhet kényelmesen az 51/82, pusztán a véletlen ingadozás miatt...



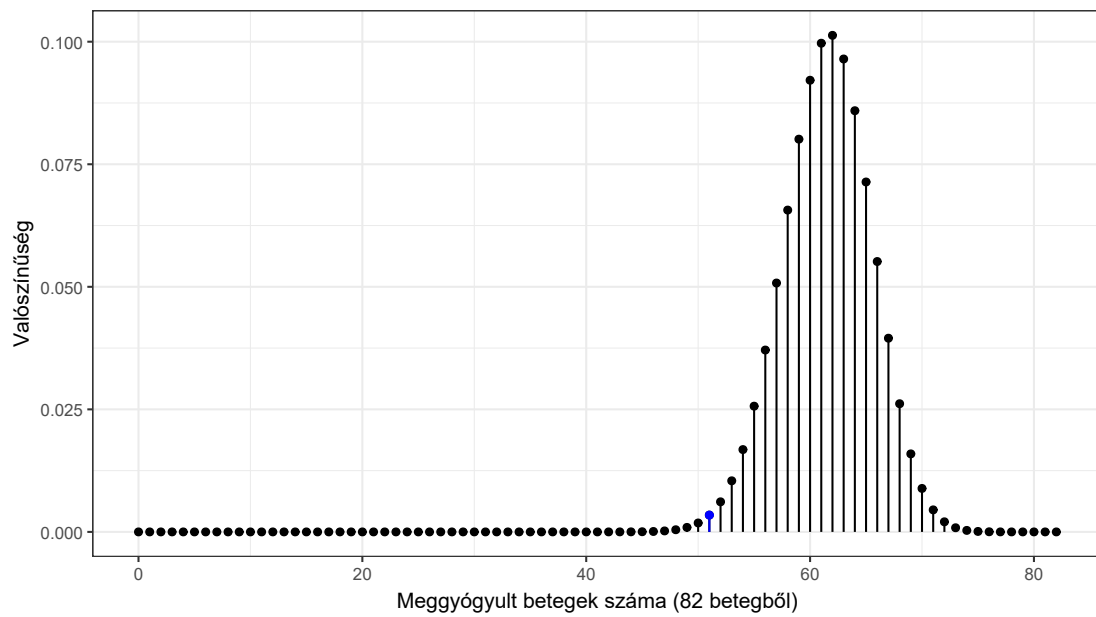
### Tippelés, mert...

...de ha a valódi gyógyulási arány 45%, akkor már nagyon valószínűtlen, hogy pusztán a véletlen ingadozás miatt 51/82 jöjjön ki



### Tippelés, mert...

...és ha a valódi gyógyulási arány 75%, akkor is nagyon valószínűtlen, hogy pusztán a véletlen ingadozás miatt 51/82 ki



### Határok húzása: a konfidenciaintervallum

- Kérdés persze, hogy mit értünk az alatt, hogy „adott valódi érték esetén kényelmesen kijöhet”
- Egy lehetséges válasz: olyan érték, amire igaz, hogy az 51/82 eltérése nem szignifikáns tőle
- Tehát, ha ezt az értéket feltételezzük valódinak, és megkérdezzük, hogy ilyen valóság esetén hihető-e (adott szignifikanciaszinten!), hogy pusztán a véletlen miatt 51/82 jöjjön ki, akkor igen a válasz
- Ezen értékek halmazát úgy hívjuk, hogy konfidenciaintervallum (CI)
- Az 51/82 (62,2%) esetén ez egy népszerű eljárással számolva: 50,8% – 72,7%
- Tehát ha a valódi arány 50,8%-nál kisebb, vagy 72,7%-nál nagyobb lenne, akkor már nem lenne hihető (adott szignifikanciaszinten!), hogy pusztán a véletlen miatt 62,2%-ot kapjunk – amit ténylegesen kaptunk...
- ...de ha a valódi érték 50,8% és 72,7% között van, akkor ezekből az értékekből kényelmesen kaphatunk 62,2%-ot a véletlen ingadozás folytán
- Vegyük észre itt is a fordított logikát: nem azt kérdeztük, hogy ha kijött 62,2%, akkor lehet-e adott érték a valóság, hanem azt, hogy ha adott érték *lenne* a valóság, akkor lehet-e 62,2% amit kapunk

### Megbízhatósági szint

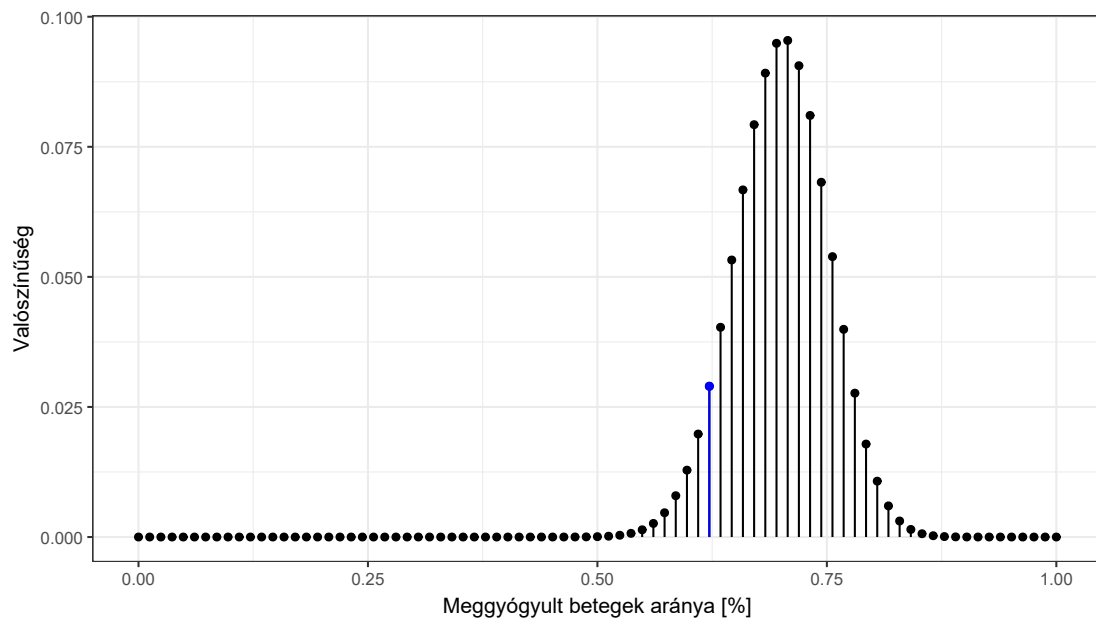
- A CI-t tehát jellemzi az az  $\alpha$  szignifikanciaszint, amit használtunk a „kényelmesen kijöhet” eldöntésére
- A gyakorlatban a CI-t inkább az  $1 - \alpha$ -val jellemezzük, neve: megbízhatósági szint
- Tehát a szokásos 5%-os szignifikanciaszint megfelelője a 95%-os megbízhatósági szint
- Egy hatás akkor nem szignifikáns 5%-on, ha a 95%-os megbízhatósági szintű CI-je tartalmazza a hatástalanságot (gondoljuk végig!)
- A CI mögött a véletlen ingadozás (még szebb nevén: mintavételi ingadozás) van: bármi a valódi érték, a mintából kapott nem mindig az lesz, hanem ingadozik körülötte – ezért van az, hogy a mintában 62,2% nem jelenti azt, hogy a valóság is annyi
- A CI szélessége jellemzi, hogy milyen precízen tudtuk „lokalizálni” a vizsgált jellemzőt (a véletlen ingadozásra tekintettel)
- A 95% egyensúlyozás: az se jó, ha azt mondjuk, hogy nagyon megbízhatóan kijelentjük, hogy szinte akárhol lehet a vizsgált jellemző, meg az sem, ha nagyon megbízhatatlanul kijelentjük, hogy szinte pontosan csak itt lehet

## A mintanagyság szerepe

- Mi történik, ha emeljük a mintanagyságot?
- Csökken a mintavételi ingadozás!
- (És így a CI is szűkebb lesz)

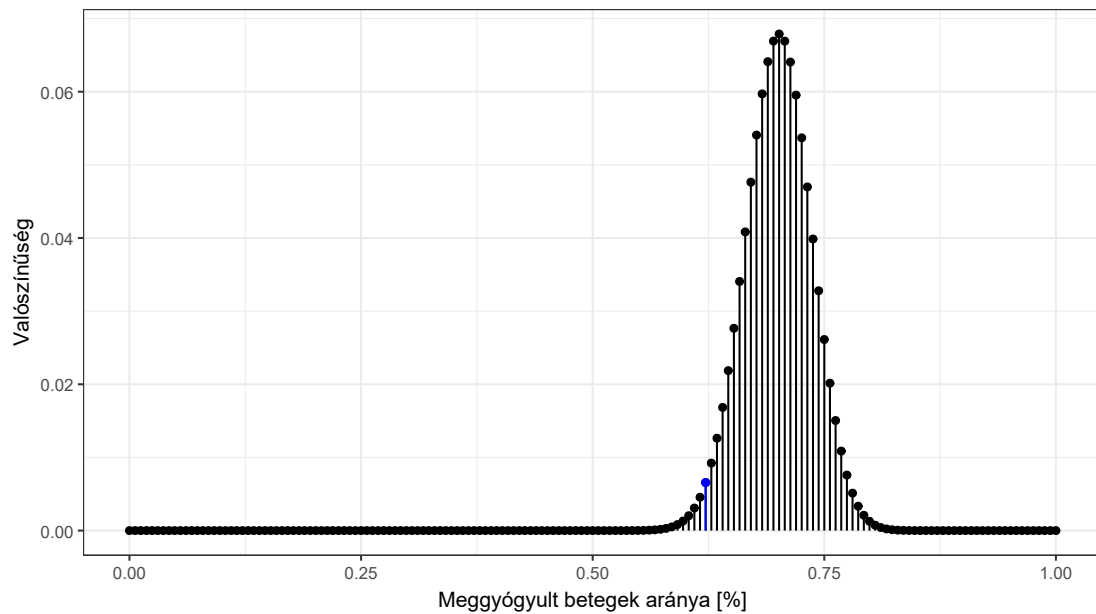
### 82 elemű minta

Ha a valóság 70% lenne, mennyire kényelmesen kaphatunk, pusztán a véletlen ingadozás miatt, 51/82-t, azaz 62,2%-ot? (95% CI: 50,8% – 72,7%)



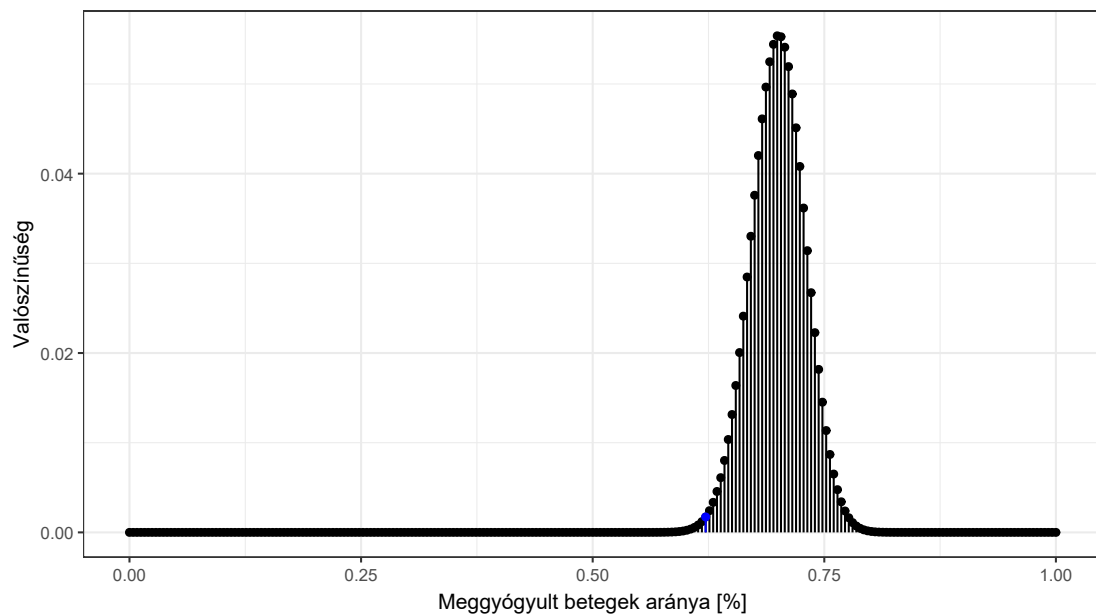
### 2 · 82 elemű minta

Ha a valóság 70% lenne, mennyire kényelmesen kaphatunk, pusztán a véletlen ingadozás miatt, 102/164-et, azaz 62,2%-ot? (95% CI: 54,3% – 69,6%)



### 3 · 82 elemű minta

Ha a valóság 70% lenne, mennyire kényelmesen kaphatunk, pusztán a véletlen ingadozás miatt, 153/246-et, azaz 62,2%-ot? (95% CI: 55,8% – 68,3%)



### Mintavételi hiba (variancia)

- A pusztán mintavételi ingadozásból fakadó hiba neve: mintavételi hiba



- A mintavételi hibára azt szokták mondani: *variancia* típusú hiba
- A becült érték eltérhet a valóságtól, de ha sokszor megismételnénk a mintavételt, akkor *átlagosan* nem lenne eltérés: ingadozik ugyan az eredményünk, de legalább a jó érték körül
- ...és minél nagyobb a mintanagyság, annál kevésbé
- Tehát ez a fajta hibázás nem bináris, nem arról van szó, hogy van ilyen vagy nincs, hanem számszerűen lemérhető: *mennyire* ingadozik a mintából kapott érték annak átlaga körül (pl. mekkora a szórása: ez az ún. standard hiba)
- Azt, hogy milyen kicsi ez a variancia szokták a vizsgálat precizitásának is nevezni
- Jól bevált apparátussal – következtető statisztika: *p*-értékek, konfidenciaintervallumok stb. – vizsgálható jellemezhető

### Nem-mintavételi hiba (torzítás)

- Minden, ami nem a mintavételi ingadozásból fakad
- ...úgyhogy egyet már nagyon is láttunk eddig is: épp ilyen a confounding is!
- Egy másik tipikus példa: mi van, ha a mintát nem olyan tökéletesen választjuk (azaz nem lehet belőle következtetni a sokaságra)?
- Általános jellemzője az ilyen hibáknak, hogy még ismételt mintavétellel sem tartanak a nullába, azaz a mintából kapott értékeknek az átlaga sem a valódi érték
- Tehát: nem is a jó érték körül ingadozik az eredmény!
- Az ilyen típusú hibákat nevezzük *torzításnak* (lemérhető az átlag és a valódi érték különbségével, ennek neve az ellenség megtévesztése végett szintén torzítás)
- Sokszor keveset beszélnek róla az orvosi gyakorlatban, pedig rettentő fontos tud lenni...

### Variancia és torzítás

- Nagyon nagy általánosságban beszélve a torzítástól félünk jobban
- Lehet a kettőt együtt is mérni: mennyire ingadozunk a jó érték körül
- Ebben benne van mindkét komponens: mennyire ingadozunk az átlag körül (variancia) és az átlag hol van a jó értékhez képest (torzítás)!
- Neve átlagos négyzetes hiba (belátható, hogy értéke a torzítás és a variancia négyzetének összege)

## Hibák vizsgálata

- Mindezek persze elméleti kérdések, olyan értelemben, hogy egyetlen minta alapján úgysem tudjuk megmondani, hogy a hibázás variancia vagy torzítás-e
- Ezért is mondtuk, hogy „képzeletbeli ismételt mintavételek” kontextusában értelmezhető
- Szebben szólva ezek vizsgálatához mind valamilyen modellt kell feltételeznünk

## A nem-mintavételi hibák szerepe

- Ki szokta úgy érezni, hogy egy adott gyógyszergyár által fizetett kísérletben mindig pont az ő gyógyszere lesz a jobb? (Pl. lelkes kutatók végignézték a 4 vezető pszichiátriai újság 1992-2002 közötti összes cikkét: amikor a kísérlet szponzora a vizsgált gyógyszer gyártója volt, akkor 78% volt a sikerarány, függetlenül 48%, kompetitor által fizetetté 28%...)
- De ez meg hogyan lehet egyáltalán? (hiszen ugyanaz a gyógyszer valakinek a saját terméke, valakinek meg a kompetitoré!)
- „Miért jobb az olanzapin mint a riszperidon, a riszperidon mint a kvetiapin és a kvetiapin mint az olanzapin?”
- Ezek tipikusan nem véletlen ingadozások, nem is direkte meghamisított eredmények, hanem a – többé vagy kevésbé jóhiszemű – nem-mintavételi hibák eredményei!

Kelly RE Jr, Cohen LJ, Semple RJ, et al. Relationship between drug company funding and outcomes of clinical psychiatric research. *Psychol Med.* 2006 Nov;36(11):1647-56. Heres S, Davis J, Maino K, et al. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am J Psychiatry.* 2006 Feb;163(2):185-94.

## A mintavételi hibák kézbentartása következtető statisztikával

- Fontos már most hangsúlyozni, hogy a mintavételi hibák kezelésére bevált apparátus ( $p$ -érték, konfidenciaintervallumok stb.) az *égegyszerre világon semmit de semmit* nem mond a nem-mintavételi hibákról!
- (Miközben simán lehet, hogy egy orvosi kutatás domináns hibaforrását ez jelenti!)

## Összefoglalva

Mintavételi hiba (elkerülhetetlen de kézbentartható, variancia típusú hiba, mintanagysággal csökken) vs. nem-mintavételi hiba (elkerülhető lehet de nehezen karakterizálható, torzítás típusú hiba, mintanagysággal nem feltétlenül csökken)