Hatásvizsgálat és kísérletezés

A címzettek nagyobb arányban vásároltak.
Vajon maga a levél is hozzájárult ehhez?

A webáruház kampánya

A példában egy webáruház 40 000 ügyfelének szimulált adataival dolgozunk.
Az ügyfelek egy része termékajánló hírlevelet kapott.

Miért mérjük a levél hatását?

- A kiküldés pénzbe kerülhet.
- A túl sok hírlevél zavarhatja az ügyfeleket.
- A levélváltozatokat először kisebb körben is kipróbálhatjuk.

Mit tudunk az ügyfelekről a kampány előtt?

- Már a kampány előtt ismertük minden ügyfél aktivitási csoportját.
- Az ügyfelek nagyjából fele az „aktívabb”, fele a „kevésbé aktív” csoportba tartozott.

Korábbi aktivitás	Ügyfelek száma	Arány a 40 000 ügyfél között
Kevésbé aktív	19 821	49,55%
Aktívabb	20 179	50,45%

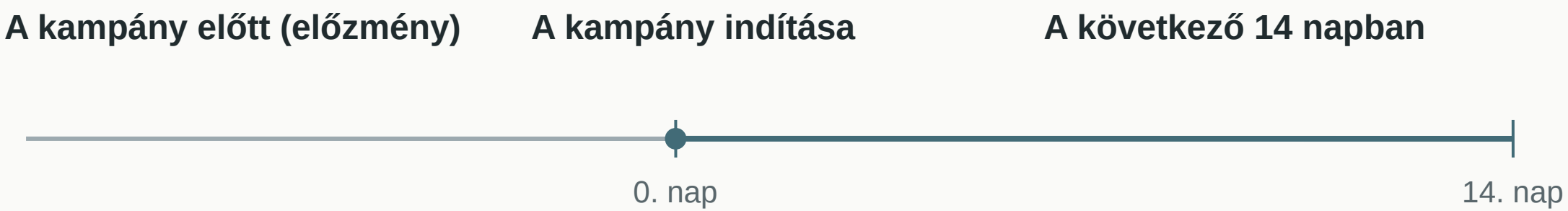
A kiküldés

A modell a korábbi vásárlások és más adatok alapján döntött. A 40 000 ügyfélből 20 159 kapott levelet (50,4%), 19 841 nem kapott.

Hány százalékponttal változna a 14 napon belül vásárlók aránya, ha mindenkinek küldenénk levelet ahhoz képest, hogy senkinek sem küldünk?

Mit mérünk?

Minden sor egy ügyfél adatait tartalmazza. Összesen 40 000 ügyfelet figyelünk meg.



A fő kimenetel a vásárlás. Minden ügyfélnél azt rögzítjük, hogy vásárolt-e a 14 nap alatt (igen/nem).

Részlet az adatokból

Hét ügyfél adatai a célzott kampányból.

Ügyfél	Korábbi aktivitás	Kapott levelet	Látogatott	Vásárolt
1	Kevésbé aktív	nem	nem	nem
2	Aktívabb	nem	igen	nem
4	Aktívabb	igen	igen	nem
5	Aktívabb	igen	igen	igen
6	Aktívabb	igen	nem	nem
8	Kevésbé aktív	igen	nem	nem
9	Kevésbé aktív	nem	igen	igen

- A webáruházban csak az vásárolhat, aki felkeresi az oldalt.
- A 2-es és a 4-es ügyfél is felkereste az oldalt, de egyikük sem vásárolt.

A levél miatt vásároltak többen?

A 14. nap végén összehasonlítjuk a levelet kapó és a levelet nem kapó ügyfelek vásárlási arányát.



Ügyfelek és vásárlók

	Csoport létszáma	Ebből vásárolt
Nem kapott levelet	19 841	1 399
Kapott levelet	20 159	3 514

Vásárlási arány = vásárlók száma / a csoport összes ügyfele

A nyers különbség: 17,43% − 7,05% = +10,38 százalékpont

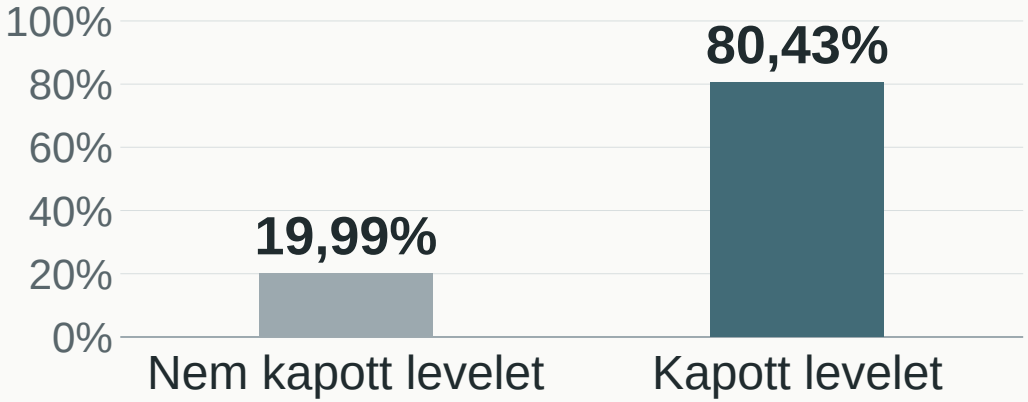
Ebből a 10,38 százalékpontos különbségből mennyit tulajdoníthatunk a levélnek?

A két csoport már a kampány előtt sem volt egyforma

Az aktívabb és a kevésbé aktív ügyfelek ugyanolyan arányban kaptak levelet?

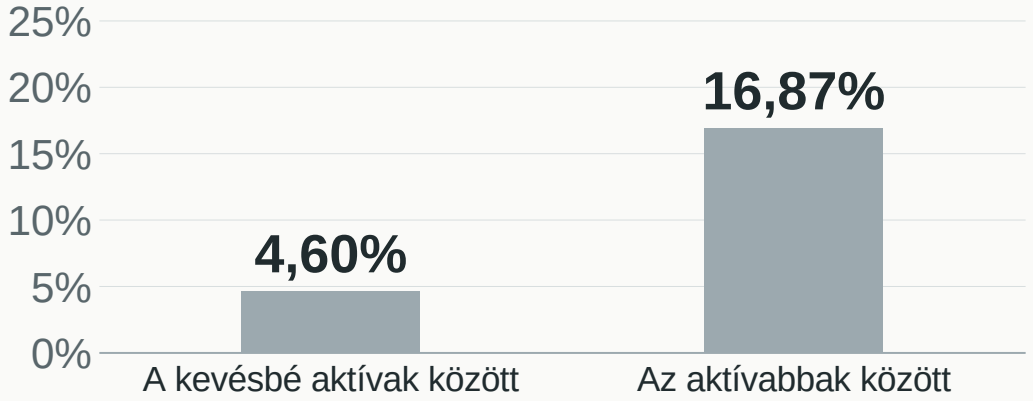
Az aktívabb ügyfelek körülbelül 80%-a kapott levelet, a kevésbé aktívaknak körülbelül 20%-a.

Kikből állt a két csoport?
A korábban aktívabb ügyfelek aránya



A címzettek 80,43%-a volt aktívabb ügyfél. A levelet nem kapók között ez az arány 19,99% volt.

A levelet nem kapó ügyfelek
hány százaléka vásárolt?



Levél nélkül az aktívabb ügyfelek 16,87%-a, a kevésbé aktívak 4,60%-a vásárolt.

- Az aktívabb ügyfelek nagyobb eséllyel kaptak levelet. Levél nélkül is nagyobb arányban vásároltak.
- A korábbi aktivitás ezért **zavaró tényező**. A +10,38 százalékpontos nyers különbség egy része abból adódik, hogy a címzettek között nagyobb az aktívabb ügyfelek aránya.

Három különböző kérdés: leírás, előrejelzés, okság

Ugyanazokból az ügyféladatokból három különböző kérdésre kereshetünk választ.

1 Leírás	Kik vásároltak nagyobb arányban? A levelet kapók 17,43%-a vásárolt, a levelet nem kapók 7,05%-a.
2 Előrejelzés	A korábbi aktivitás alapján kiknél valószínűbb a vásárlás? Egy modell a becsült vásárlási esély szerint rangsorolhatja az ügyfeleket.
3 Oksági kérdés	Hogyan változna a vásárlás esélye, ha ugyanezeknek az ügyfeleknek levelet küldenénk? Mi történne velük levél nélkül?

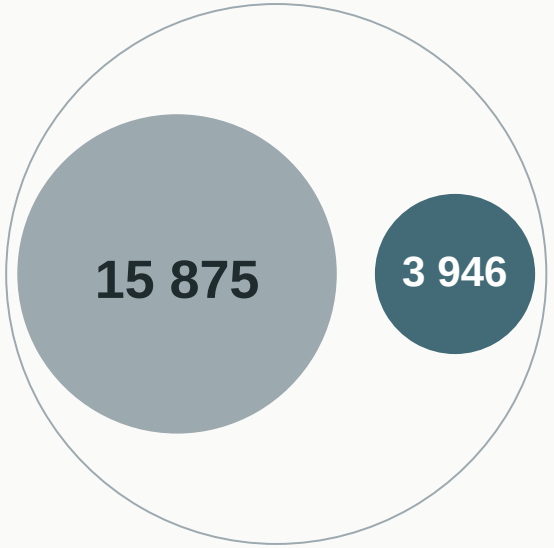
Attól, hogy valaki nagy eséllyel vásárol, még nem tudjuk, változtatna-e a döntésén a hírlevél.

Kiigazítás (1. lépés): összehasonlítás azonos aktivitás mellett

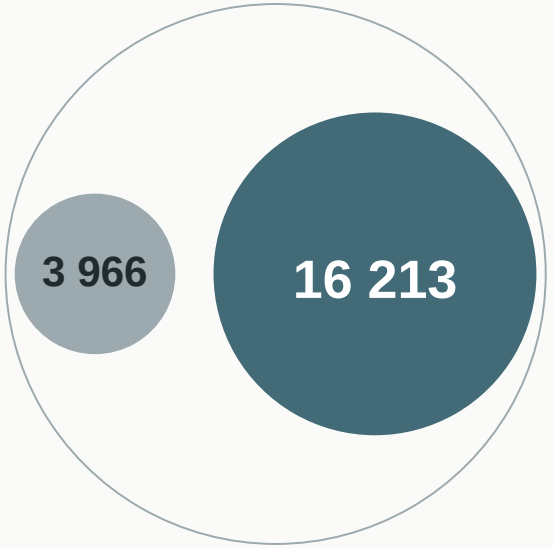
Egy ügyfélnél a két lehetőség közül csak az egyiket látjuk. Ha például kapott levelet, nem tudjuk megfigyelni, hogy levél nélkül is vásárolt volna-e. Ez a hiányzó ellenpont.

A célzott kampány 40 000 ügyfelét a kampány előtti aktivitásuk szerint két csoportra bontjuk.

Korábban kevésbé aktív (19 821 fő)



Korábban aktívabb (20 179 fő)



Nem kapott levelet Kapott levelet

Mindkét aktivitási csoporton belül a levelet kapók vásárlási arányát hasonlítjuk a levelet nem kapókéhoz.

Kiigazítás (2. lépés): +10,38 helyett +2,34 százalékpont

A két vásárlási arány különbségét százalékpontban (szp.) adjuk meg.

1 A két vásárlási arány különbsége az egyes aktivitási csoportokban

Korábbi aktivitás	Nem kapott levelet	Kapott levelet	Különbség
Kevésbé aktív	730 / 15 875 (4,60%)	233 / 3 946 (5,90%)	+1,31 szp.
Aktívabb	669 / 3 966 (16,87%)	3 281 / 16 213 (20,24%)	+3,37 szp.

2 Átlag egy 1 000 fős ügyfélkörre

Vegyünk 1 000 ügyfelet: 500 kevésbé aktívat és 500 aktívabbat. Így mindkét csoport súlya pontosan ½.

½ × 1,31 szp. + ½ × 3,37 szp. = +2,34 százalékpont

Ha ennek az 1 000 ügyfélnek levelet küldünk, a becsült többlet átlagosan 23,4 vásárló a levél nélküli esethez képest.

Mikor elég az utólagos csoportosítás, és mikor nem?

A célzott kampányban az aktivitási csoportokon belül hasonlítottuk össze az ügyfeleket. Így kaptuk a +2,34 százalékpontos becslést.

Mikor értelmezhetjük a kiigazított különbséget a levél hatásaként?

Ha az aktivitási csoportokon belül már nincs olyan tényező, amely a kiküldést és a vásárlási esélyt is befolyásolja. Mindkét aktivitási csoportban kellenek levelet kapó és levelet nem kapó ügyfelek.

Mi nehezítheti az összehasonlítást?

1. Nem mért közös okok. Ha a pillanatnyi vásárlási szándék a címzettek kiválasztását és a vásárlást is befolyásolta, az aktivitás szerinti kiigazítás után is maradhat torzítás.

2. Kevés ügyfél az alcsoportokban. Már 10 kétértékű jellemzőből is 1 024 alcsoport képezhető. Lehet olyan kis elemszámú alcsoport, ahol mindenki kapott levelet, így nincs kivel összehasonlítani a címzetteket (mert nincs olyan aki nem kapott).

Mi lett volna, ha ugyanebben a kampányban sorsolással választjuk ki a címzetteket?

Ki kap levelet az A/B-tesztben?

- Ugyanaz a 40 000 ügyfél, ugyanaz a hírlevél.
- Minden ügyfélnél külön sorsolunk. A levélküldés esélye 50%.
- Mindkét csoportban megnézzük, hányan vásároltak 14 napon belül.

A sorsolásos besorolás neve randomizálás. Az A csoport tagjai nem kapnak levelet (kontrollcsoport), a B csoport tagjai kapnak (kezelt csoport). Ilyen A/B-tesztet kisebb mintán is végezhetünk.

	Teljes ügyfélkör	Nem kapott levelet (A)	Kapott levelet (B)
Ügyfelek száma	40 000	20 196	19 804
Ebből aktívabb ügyfél	20 179	10 114	10 065
Aktívabbak aránya a csoportban	50,45%	50,08%	50,82%

Az A csoportban 20 196 ügyfélből 10 114 volt korábban aktívabb. Az arányuk 50,08%.

A sorsolás mindenkinek azonos esélyt ad arra, hogy levelet kapjon. A csoportok összetétele kissé eltérhet. Az aktívabbak aránya az A csoportban 50,08%, a B csoportban 50,82% volt.

Az A/B-tesztben mért eredmény: +2,75 százalékpont

Sorsolással döntöttük el, ki kap levelet. Ezért a levél átlagos hatását közvetlenül a két csoport 14 napos vásárlási arányának különbségével becsülhetjük.

Csoport	Csoport létszáma	Ebből vásárolt	Vásárlási arány
Nem kapott levelet (A)	20 196 fő	2 150 fő	2 150 / 20 196 = 10,65%
Kapott levelet (B)	19 804 fő	2 653 fő	2 653 / 19 804 = 13,40%

Az A/B-tesztben mért különbség: 13,40% − 10,65% = +2,75 százalékpont

1 000 hasonló ügyfélre vetítve a becsült többlet körülbelül 27–28 vásárló.

Ugyanarra a levélre a kiigazítás +2,34, az A/B-teszt +2,75 százalékpontos becslést adott.
Mekkora a valódi átlagos hatás, és miért térhet el tőle a két becslés?

Mennyi a valódi hatás, és miért térnek el tőle a méréseink?

A szimuláció szabályaiból mindkét aktivitási csoportban kiszámíthatjuk a vásárlás esélyét levéllel és levél nélkül. A két esély különbsége megadja a levél hatását az adott csoportban.

A vásárlás valószínűsége és a levél hatása a két aktivitási csoportban

Korábbi aktivitás	Vásárlás esélye levél nélkül	Vásárlás esélye levéllel	Valódi hatás
Kevésbé aktív	4,8%	6,4%	+1,6 szp.
Aktívabb	16,2%	19,6%	+3,4 szp.

A levél átlagos hatása a két csoport azonos súlyával 0,5 × 1,6 szp. + 0,5 × 3,4 szp. = +2,50 százalékpont

A modell által célzott kampány adataiból kiigazítás után +2,34 százalékpontra becsültük a levél átlagos hatását.
Az A/B-tesztben +2,75 százalékpontot mértünk.
Egy kísérlet eredménye attól is függ, kik kerülnek a két csoportba, és közülük kik vásárolnak.

Mennyit ingadozhat az A/B-teszt becslése 40 000 fős mintákban?

Miért térhet el a becslés kísérletről kísérletre?

- Minden szimulált A/B-teszthez új, 40 000 fős ügyfélmintát veszünk.
- A hírlevél minden tesztben ugyanaz. Minden ügyfélnél sorsolás dönt a kiküldésről, 50% eséllyel kap levelet.

Mi marad állandó?

A modellben a levél valódi átlagos hatása minden ismétlésnél +2,50 százalékpont.

Mi változhat?

Ismétlésenként más ügyfelek kerülnek a mintába. Változik, ki kap levelet, és hányan vásárolnak a két csoportban.

Az első kísérlet becslése és öt példa arra, milyen értéket kaphatunk ismétléskor

1. kísérlet (a mi mérésünk)	2. kísérlet	3. kísérlet	4. kísérlet	5. kísérlet	6. kísérlet
+2,75 szp.	**+2,21 szp.**	**+2,63 szp.**	**+2,92 szp.**	**+2,38 szp.**	**+3,10 szp.**

Egy kísérletből egy pontbecslést kapunk. Az első kísérletben ez +2,75 százalékpont volt.

Sok kísérlet pontbecslései a modellben rögzített +2,50 százalékpont körül szóródnak.

Hogyan fejezhetjük ki egyetlen kísérlet adataiból a becslés bizonytalanságát?

Mekkora a +2,75 százalékpontos becslés standard hibája?

Általános összefüggés

$$\mathrm{SE}(\hat{\theta}) = \sqrt{\mathrm{Var}(\hat{\theta})}$$

A $\hat{\theta}$ a mintából számolt becslést jelöli. A standard hiba azt mutatja meg, mekkora ennek a becslésnek a szórása, ha újra és újra mintát veszünk.

Két független csoport vásárlási arányának különbsége

A két arány különbségének varianciája a függetlenség miatt a két variancia összege. Ezért van pluszjel a gyök alatt.

Levelet kaptak (B) 2 653 / 19 804 ≈ 0,1340 **Nem kaptak levelet (A)** 2 150 / 20 196 ≈ 0,1065

Képlet

$$SE = 100 \times \sqrt{\frac{\hat{p}_B \times (1 - \hat{p}_B)}{n_B} + \frac{\hat{p}_A \times (1 - \hat{p}_A)}{n_A}}$$

Behelyettesítés

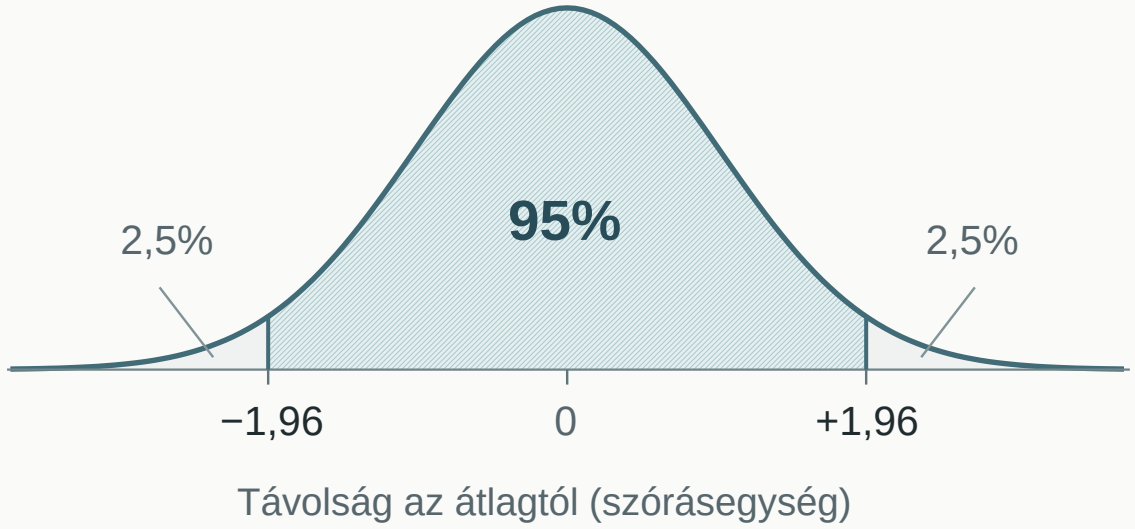
$$SE = 100 \times \sqrt{\frac{0{,}1340 \times (1 - 0{,}1340)}{19\,804} + \frac{0{,}1065 \times (1 - 0{,}1065)}{20\,196}} \approx 0{,}325 \text{ százalékpont}$$

p̂ a csoportban megfigyelt vásárlási arány, n a csoport létszáma. A 100-as szorzó százalékpontra vált.

Hogyan kapjuk meg a 95%-os konfidenciaintervallumot?

A mért különbség +2,75 százalékpont. A becsült standard hiba (SE) 0,325 százalékpont.

Standard normáleloszlás



Miért 1,96?

A görbe alatti terület 95%-a −1,96 és +1,96 közé esik. A két szélen 2,5–2,5% marad.

95%-os hibahatár normálközelítéssel

1,96 × 0,325 ≈ 0,64 százalékpont

95%-os konfidenciaintervallum

2,75 ± 0,64 szp., azaz +2,11-től +3,39 százalékpontig.

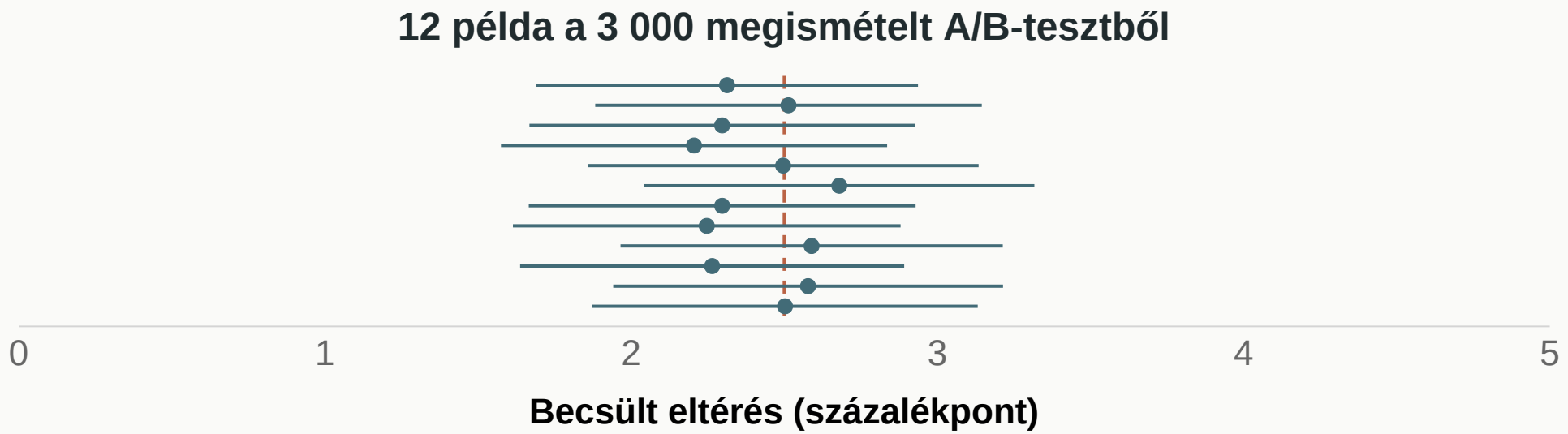
A teljes intervallum nulla fölött van. Az adatok alátámasztják, hogy a levél növelte a vásárlás esélyét.

Ha sokszor megismételnénk a kísérletet, az intervallumok kb. 95%-a tartalmazná a valódi hatást.

Mit jelent a 95%-os lefedettség sok megismétlés esetén?

Ugyanazzal a modellel 3 000 független, egyenként 40 000 fős A/B-tesztet szimuláltunk. Az ábra 12 kísérlet eredményét mutatja.

Minden sor egy kísérlet. A pont a becsült hatás, a vízszintes vonal az adott kísérlet 95%-os konfidenciaintervalluma. Ha a vonal metszi a függőleges szaggatott vonalat, az intervallum tartalmazza a valódi hatást (+2,50 szp.).



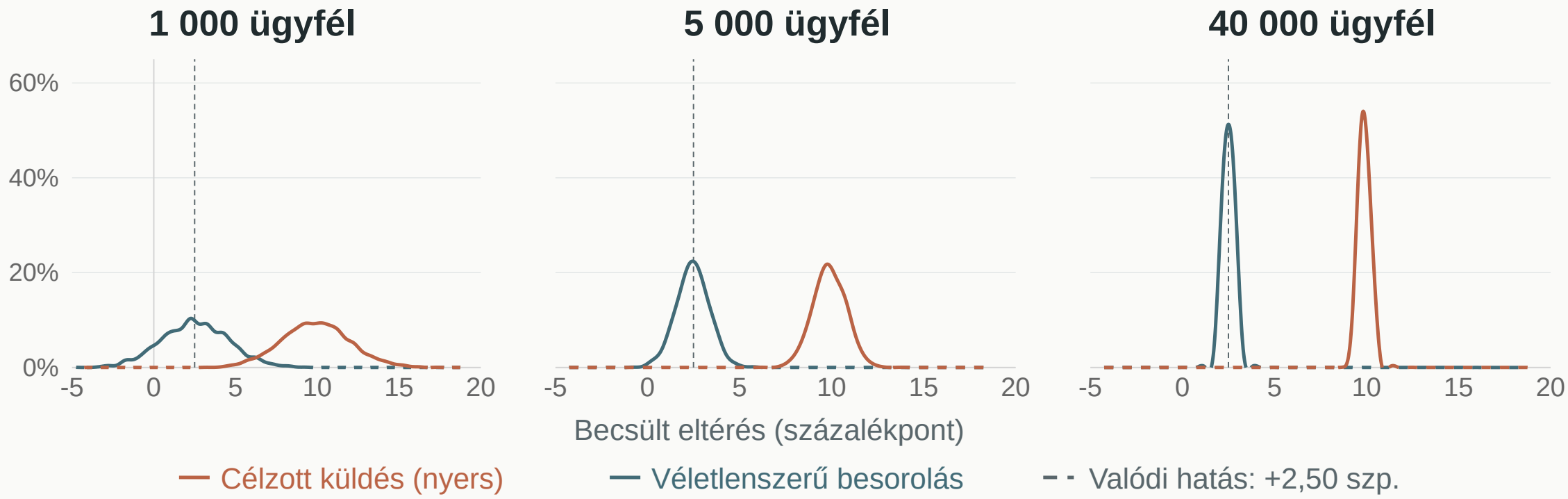
A bemutatott 12 intervallum mindegyike tartalmazza a +2,50 százalékpontot.

Mind a 3 000 kísérletet nézve az intervallumok 94,6%-a tartalmazta a valódi hatást. Valós kampánynál a valódi hatás ismeretlen, így egyetlen intervallumról ezt nem tudjuk ellenőrizni.

Mit javít meg a nagyobb minta, és mit nem?

A görbék azt mutatják, az ismételt szimulációk mekkora részében esett egy-egy tartományba a vásárlási arányok különbsége. Nagyobb mintánál a becslések kevésbé szóródnak.

Az A/B-tesztben a becsült különbség közelítő 95%-os hibahatára 1 000 ügyfélnél ±4,1 szp., 5 000-nél ±1,8 szp., 40 000-nél pedig ±0,64 szp.

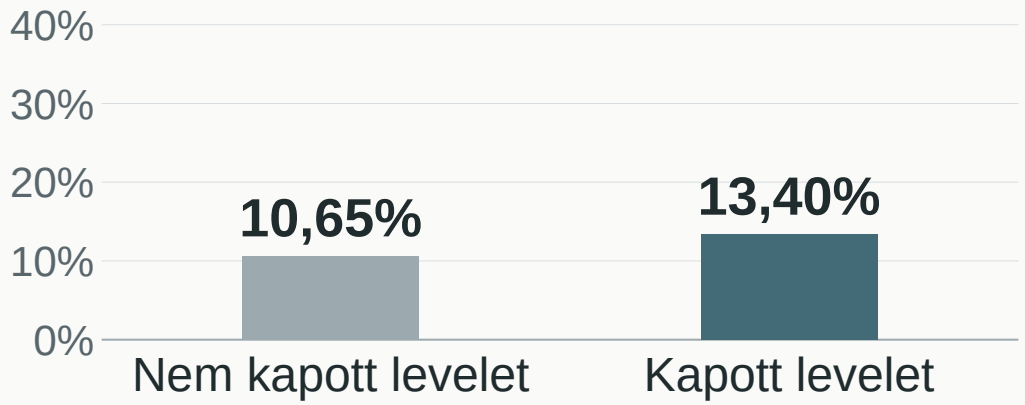


Az A/B-teszt becslései a valódi +2,50 százalékpont körül sűrűsödnek. A célzott küldés nyers különbségei nagy mintánál is +10 százalékpont körül maradnak.

Egy gyakori csapda: miért ne szűrjünk csak a látogatókra?

A webáruházban csak az vásárolhatott, aki meglátogatta az oldalt. Hogyan változik a két csoport összehasonlítása, ha csak a látogatókat vesszük figyelembe?

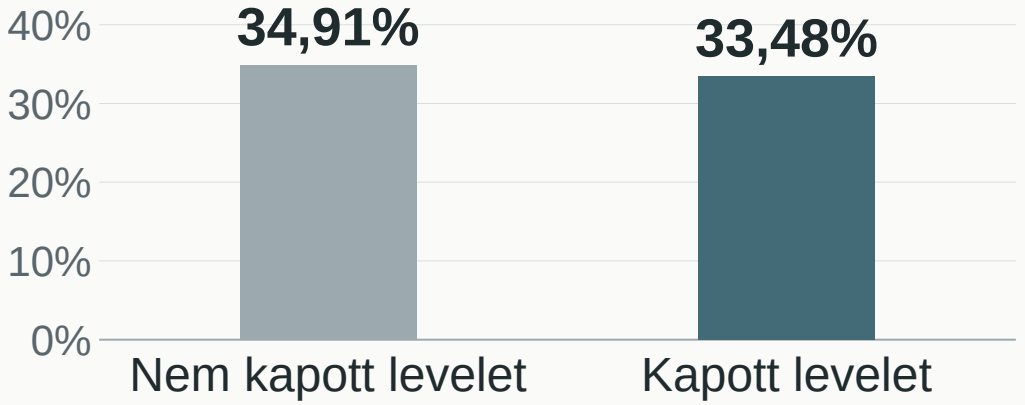
Az összes besorolt ügyfél hány százaléka vásárolt?



Levél nélkül 2 150 vásárló / 20 196 ügyfél (10,65%)
Levéllel 2 653 vásárló / 19 804 ügyfél (13,40%)

Különbség a teljes mintán: +2,75 százalékpont

A látogatók hány százaléka vásárolt?



Levél nélkül 2 150 vásárló / 6 158 látogató (34,91%)
Levéllel 2 653 vásárló / 7 923 látogató (33,48%)

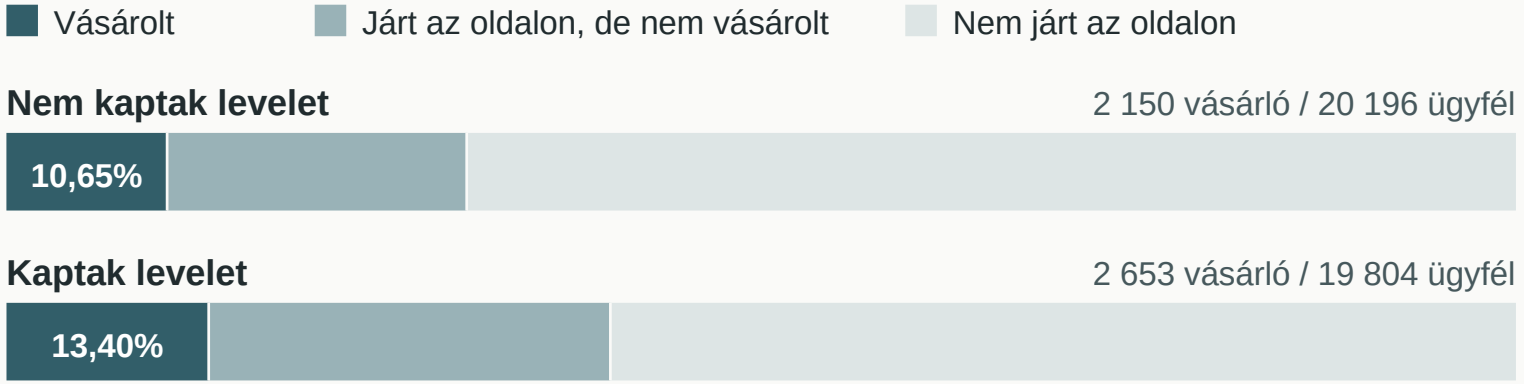
Különbség a látogatók között: −1,43 százalékpont

Miért pozitív a különbség az összes ügyfél között, és miért negatív, ha csak a látogatókat nézzük?

A levél azt is befolyásolja, ki kerül a látogatók közé

Ugyanazokat a vásárlókat számoljuk össze. Az eredmény attól függ, hogy az összes ügyfélhez vagy csak a látogatókhoz viszonyítjuk a számukat.

Minden sáv az adott csoport összes ügyfelét mutatja



Az összes ügyfélnél +2,75 százalékpont

Csak a látogatók között

Nem kaptak levelet

6 158 látogató (30,5%)
2 150 / 6 158 = 34,91%

Kaptak levelet

7 923 látogató (40,0%)
2 653 / 7 923 = 33,48%

A látogatóknál −1,43 szp.

- A levelet kapók közül jóval többen látogattak el az oldalra, így végül több volt a vásárló, pedig a látogatók kisebb része vásárolt.
- A levél olyanokat is az oldalra hozhatott, akik csak körülnéztek.

Ez kiválasztási torzításhoz vezethet: maga a levél is befolyásolja, ki kerül a vizsgált látogatók közé.

Mit mond a látogatók közötti különbség?

A levelet kapó látogatók vásárlási aránya 1,43 százalékponttal alacsonyabb volt. Mennyire pontos ez a különbség, és mit mond a levél hatásáról?



Összeegyeztethető a látogatók adataival a nulla különbség?
Igen. A becslés −1,43 szp. A 95%-os intervallum −3,01-től +0,15 szp.-ig terjed, így a nullát is tartalmazza.

Mit hasonlítsunk össze a levél teljes hatásának becsléséhez?
A levelet kapók és a levelet nem kapók vásárlási arányát. Mindkét arányt a csoport teljes létszámával számoljuk.

Több látogató adatából a látogatók közötti különbséget becsülhetnénk pontosabban.

Megéri elküldeni 1 000 levelet?

Az A/B-tesztben a becsült hatás +2,75 százalékpont. A 95%-os konfidenciaintervallum +2,11-től +3,39 százalékpontig terjed.

Tegyük fel, hogy a bevételből a termék és a kiszolgálás költségét levonva vásárlónként átlagosan 20 € marad. Becslésünk szerint 1 000 levél hatására 27–28-cal többen vásárolnak. Ez körülbelül 550 €-t hoz a webáruháznak. Ebből még levonjuk a levélküldés költségét.

Becsült nyereség = 27,5 × 20 € − 1 000 × egy levél küldési költsége

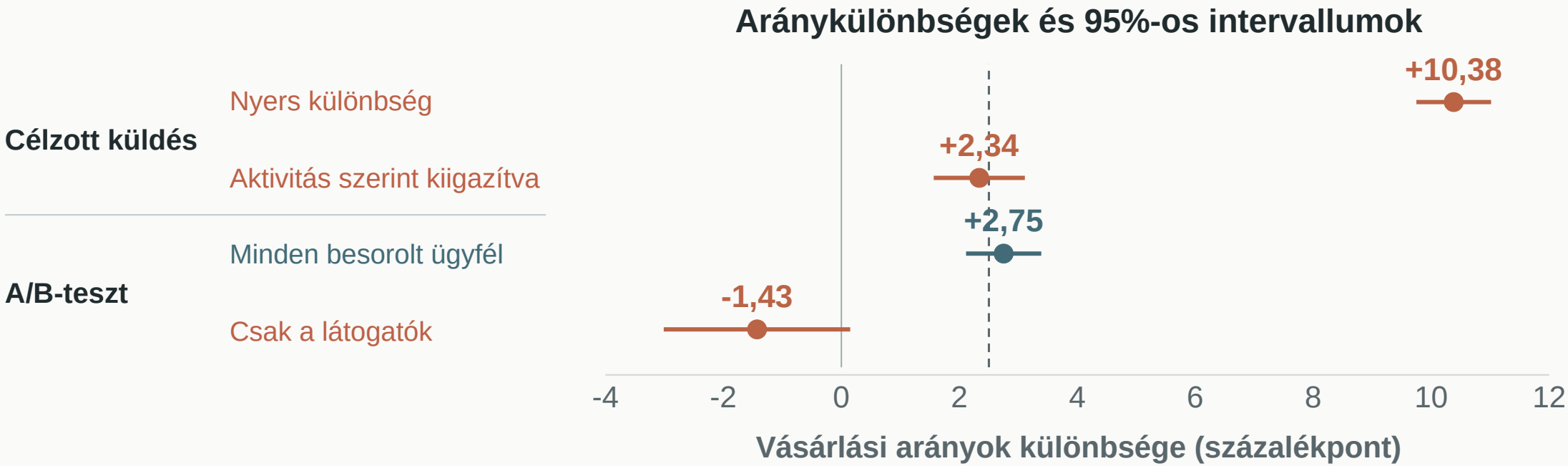
Mennyivel kell nőnie a vásárlók arányának ahhoz, hogy megtérüljön a küldés?

Levélköltség	**Megtérülési küszöb** (egy levél küldési költsége / 20 €)	**Hol van a megtérülési küszöb az intervallumhoz képest?**
0,20 € / levél	1,00 szp. kell a nullszaldóhoz	A teljes intervallum az 1 szp.-os küszöb fölött van. Az adatok alátámasztják a nyereségességet.
0,60 € / levél	3,00 szp. kell a nullszaldóhoz	A 3 szp.-os küszöb az intervallumba esik. A becsült eredmény 1 000 levélnél 550 − 600 = −50 €. Nyereség és veszteség is lehetséges.
0,80 € / levél	4,00 szp. kell a nullszaldóhoz	A teljes intervallum a 4 szp.-os küszöb alatt van. Ekkora küldési költségnél veszteség várható.

A becsült hatás mindhárom esetben ugyanaz. A döntéshez az intervallumot az adott levélköltséghez tartozó megtérülési küszöbhöz viszonyítjuk.

Összegzés: négy összehasonlítás ugyanazon az ügyfélkörön

Ugyanahhoz a 40 000 ügyfélhez kétféle kiküldést szimuláltunk. A célzott kampányt kiigazítás előtt és után értékeltük. Az A/B-tesztben először minden besorolt ügyfelet, majd csak a látogatókat vizsgáltuk.



A nyers célzott összehasonlítás (+10,38 szp.) és a látogatókra szűrés (−1,43 szp.) eltérő összetételű csoportokra épül.

Az A/B-teszt becslése +2,75 szp. A kiigazított +2,34 szp. is a levél hatását becsüli, mert itt a célzás csak az aktivitástól függött.

A szaggatott vonal a modellben rögzített valódi átlagos hatást jelöli (+2,50 százalékpont).

Hogyan vizsgáljuk egy beavatkozás hatását?

Mit mérünk?	Előre rögzítsük, kiknél, mihez képest, milyen eredményt és milyen időtávon vizsgálunk.
Hogyan alakult ki a két csoport?	A/B-tesztben sorsolással döntünk a besorolásról. Célzott küldésnél kiigazításra van szükség, ha egy korábbi jellemző a kiválasztást és a vásárlást is befolyásolja.
Kiket számítunk bele?	A teljes hatás becsléséhez mindkét kisorsolt csoport minden tagját vegyük figyelembe. A levél például azt is befolyásolja, ki látogat el az oldalra.
Mekkora a hatás, és mennyire pontos a becslés?	A két vásárlási arány különbségét százalékpontban adjuk meg. A standard hiba a becslés kísérletről kísérletre várható ingadozását jellemzi.
Mit mutat a 95%-os konfidenciaintervallum?	Ha a különbség intervalluma tartalmazza a nullát, az adatokkal a nulla különbség is összeegyeztethető. A megtérüléshez nézzük meg, hogy az intervallum a küszöb fölött van, alatta van, vagy tartalmazza a küszöböt.

Rövid gondolatkísérlet: vizsgafelkészítő kurzus

A vizsgafelkészítőre a hallgatók önként jelentkeznek.

A résztvevők jobb vizsgaeredményt értek el. Mennyit tulajdoníthatunk ebből a felkészítőnek?

1 **Önkiválasztás és kezdeti különbségek**

A korábbi tudás már a kurzus előtt eltérhetett. Befolyásolhatta azt is, hogy ki jelentkezik a felkészítőre, és azt is, hogy ki hogyan teljesít a vizsgán.

2 **Milyen összehasonlítás lenne hiteles?**

Sorsoláskor a bejutókat a ki nem sorsolt jelentkezőkkel vetnénk össze. Így a kurzushely felajánlásának hatását becsülnénk. A korábbi jegyek szerinti kiigazításnál más, nem mért különbségek is számíthatnak.

3 **Segítene-e a több adat?**

Ha 100 helyett 10 000 hallgatót vizsgálunk, pontosabban becsülhetjük a megfigyelt különbséget. Az önkéntes jelentkezésből eredő torzítás ettől megmaradhat.

A kurzus hatásának megítéléséhez azt is tudnunk kell, hogyan kerültek a hallgatók a két csoportba.