

Statistická validita CX dat.

Kolik odpovědí opravdu potřebujete?

Obsah

01	Úvod	3
02	Tři síly, které určují váš minimální vzorek	4
03	Referenční tabulka velikosti vzorku	6
04	Spolehlivost vs. přesnost — trade-off	8
05	Response-rate trychtýř	10
06	Segmentace dat — velikost vzorků pro segmenty	12
07	Čtení pohybů — kdy je změna skutečná?	14
08	Šest statistických chyb, které CX týmy dělají	16
09	Sebehodnotící scorecard	18
—	Závěrečné myšlenky	19

Úvod.

Způsob, jakým CX týmy zacházejí se svými daty, ovládají dva typy selhání — a jsou protikladné. Jeden tým reportuje 0,4bodový posun NPS exekutivnímu výboru na vzorku 38 odpovědí a doporučuje akci. Druhý tým s 1 200 odpověďmi odmítá reagovat, protože výsledek „není dostatečně významný“. Oba se mýlí a oba jsou běžné.

Statistická gramotnost v CX není pokročilá statistika — je to vědět, jaký minimální vzorek potřebujete, vědět, jak číst pohyb, a vědět, kdy přestat počítat.

Tento dokument pokrývá pracovní znalosti, které řeší oba typy selhání. Tři síly určují váš minimální vzorek — populace, spolehlivost a chyba odhadu. Jedna referenční tabulka odpoví na „kolik jich potřebuju?“ pro téměř každý běžný CX scénář. Dvě další kapitoly pokrývají plánování response rate a segmentaci — místa, kde většina CX programů neúmyslně sbírá málo. Poslední technická kapitola pokrývá čtení pohybu: kdy je změna skutečná a kdy je to šum.

Nic z tohoto nevyžaduje titul ze statistiky. Cílem dokumentu je dát CX manažerovi dost znalostí, aby ubránil rozhodnutí o velikosti vzorku před CFO — a dost skepse, aby zatlačil zpět, když exekutiva požaduje akci na vzorku čtyřiceti odpovědí.

CO SE DOZVÍTE

Tři síly, které určují jakýkoliv minimální vzorek.

Referenční tabulku pokrývající populace od 100 do 1 000 000.

Vzorec, jak zpětně dopočítat počet pozvánek z response rate.

Upravené velikosti vzorků pro analýzu na úrovni segmentu.

Pravidlo pro čtení pohybů mezi obdobími bez falešných poplachů.

Šest běžných statistických chyb — a jak se každé vyhnout.

PRO KOHO TO JE

CX ředitele obhajující velikosti vzorků před vedením.

VoC programové leady navrhující transakční průzkumy.

Analytiky čtoucí dashboardy a rozhodující, co je signál a co šum.

Každého, koho někdy zaskočila otázka „je to statisticky významné?“

CO TADY NENAJDETE

Hlubokou statistickou teorii ani odvození vzorců. Cílem je dát vám pracovní pravidla, která ubráníte u tabule — ne vás naučit, jak se k nim došlo.

KAPITOLA 1

Tři síly.

Spolehlivost, chyba odhadu, populace — a jedna běžná mylná představa.

1

Tři „knoflíky“ — a jedna běžná mylná představa.

Téměř každá otázka „jak velký má být vzorek?“ se redukuje na tři nastavení. Žádné z nich není velikost vaší populace — což je nejčastější mylná představa v CX. Požadovaná velikost vzorku se po překročení několika tisíc populace téměř nemění.

01

Úroveň spolehlivosti.

Jak jistí chcete být, že váš výsledek odráží realitu? CX standard je 95 % — to znamená, že kdybyste průzkum spustili 100krát, skutečná hodnota by byla uvnitř vašeho intervalu v 95 z nich. 90 % stačí na rané skenování; 99 % je téměř na všechno overkill.

02

Chyba odhadu (margin of error).

Jak přesní potřebujete být? ± 5 procentních bodů je CX default a funguje na trackování. ± 3 body jsou potřeba, když rozhodnutí pohne penězi. Posun pod ± 3 body zhruba ztrojnásobuje velikost vzorku za každý 1bodový posun přesnosti.

03

Velikost populace.

Kontraintuitivně záleží míň, než si lidi myslí. Populace 10 000 potřebuje téměř stejný vzorek jako populace 1 000 000. Pod $\sim 1\,000$ záleží; nad $\sim 10\,000$ můžete obvykle ignorovat.

SKRYTÁ ČTVRTÁ PROMĚNNÁ: OČEKÁVANÉ ROZDĚLENÍ

Vzorce předpokládají nejhorší možné rozdělení (50/50). Pokud z dřívějších dat víte, že výsledek je blíž 80/20 (např. 80 % promotérů), můžete se obejít s menšími vzorky. V praxi CX týmy používají 50/50 předpoklad — je bezpečnější a matematika je čistší. Upravujte jen, když máte stabilní historická data, o která se můžete opřít.

PRACOVNÍ DEFAULTY

Pro 95 % spolehlivost a ± 5 procentních bodů na populaci nad 10 000 potřebujete **385 odpovědí**.

Pro ± 3 body při stejné spolehlivosti potřebujete zhruba **1 070 odpovědí**.

Tahle dvě čísla si zapamatujte — pokrývají většinu CX use case'ů.

KAPITOLA 2

Referenční tabulka.

Velikost vzorku podle populace a požadované přesnosti.

2

Velikost vzorku podle populace a přesnosti.

Tabulka níže je hlavní artefakt tohoto dokumentu. Najděte si řádek odpovídající vaší populaci, sloupec odpovídající vaší požadované přesnosti, a číslo v buňce je minimální počet **dokončených odpovědí** — ne rozeslaných pozvánek — pro obhajobu 95% intervalu spolehlivosti.

POPULACE (N)	±5 % · 90 % SPOL.	±5 % · 95 % SPOL.	±3 % · 95 % SPOL.	±3 % · 99 % SPOL.	POZNÁMKY
100	74	80	92	99	Skoro každý
500	176	218	341	446	Většina B2B účtů
1 000	214	279	517	747	Boutique B2C
5 000	258	358	881	1 485	Mid-tier retail
10 000	264	371	965	1 719	Běžný strop
50 000	270	382	1 046	1 945	Téměř totožné
100 000	270	384	1 057	1 975	s 10 000
1 000 000+	271	385	1 068	2 000	Asymptota

JAK ČÍST TUHLE TABULKU

Zvýrazněný sloupec (95 % spolehlivost, ±5 % chyba odhadu) je CX default — sáhněte po něm, pokud nemáte konkrétní důvod jinak. Všimněte si té asymptoty: jakmile vaše populace překročí ~10 000, požadovaný vzorek se téměř nehne. To je jediný nejužitečnější fakt v tomto dokumentu.

KAPITOLA 3

Spolehlivost vs. přesnost.

Trade-off, kde většina týmů přepáluje spolehlivost a podstřelí přesnost.

3

Jak vybrat správnou spolehlivost a chybu odhadu.

Ty dva „knoflíky“ spolu interagují nelineárně. Posun z 90 % na 95 % spolehlivosti zvyšuje požadovaný vzorek zhruba o 40 %. Posun z ± 5 % na ± 3 % ho zvyšuje zhruba o 280 %. Většina týmů přepaluje spolehlivost a podstřeluje přesnost — přitom u většiny CX rozhodnutí záleží víc přesnost.

Kdy použít kterou kombinaci

90 %, ± 5 %

Rané programy nebo směrové čtení.

Rychlé měsíční skenování. Zhruba 271 odpovědí pro jakoukoliv velkou populaci.

95 %, ± 5 %

Default pro běžné CX trackování.

Číslo, které byste měli umět obhájit u jakéhokoliv exekutivního reviewu. 385 odpovědí pro jakoukoliv velkou populaci.

95 %, ± 3 %

Když výsledek pohne investicemi.

Nebo když pohne performance reviews. 1 068 odpovědí pro jakoukoliv velkou populaci.

Kolik vás to stojí

NASTAVENÍ	POŽADOVANÝ VZOREK	VS. BASELINE
90 % spolehlivost, ± 5 %	271	0,70x
95 % spolehlivost, ± 5 %	385	1,00x (baseline)
95 % spolehlivost, ± 3 %	1 068	2,77x
99 % spolehlivost, ± 5 %	666	1,73x
99 % spolehlivost, ± 3 %	1 849	4,80x
99 % spolehlivost, ± 1 %	16 641	43x

CO TO ZNAMENÁ

Honit 99 % spolehlivost při ± 1 % vyžaduje 43x vzorek defaultu — pro zisk přesnosti, který žádné reálné CX rozhodnutí nepotřebuje.

KAPITOLA 4

Response-rate trychtýř.

Kolik pozvánek poslat, abyste dostali ten vzorek, který potřebujete.

4

Zpětně z dokončených odpovědí.

Velikost vzorku z kapitoly 2 je počet **dokončených odpovědí** — ne počet pozvánek, které pošlete. Abyste věděli, kolik pozvánek potřebujete, potřebujete povlivý odhad svého response rate, který je téměř vždy nižší, než CX týmy očekávají.

PRACOVNÍ VZOREC

$$\text{Pozvánky} = \text{Požadované odpovědi} \div \text{Response rate}$$

Typické response raty

Cold email průzkum	3–8 %
Warm email (stávající zákaznický list)	5–15 %
Post-transaction email prompt	15–30 %
In-product widget	20–40 %
B2B s osobním kontaktem	30–60 %
Telefonní (CATI) průzkum	20–35 %

ŘEŠENÝ PŘÍKLAD

Předpokládejme, že potřebujete 385 dokončených odpovědí (95 % spolehlivost, ± 5 %, velká populace) a používáte post-transaction email prompt s očekávaným response rate 18 %.

$$\text{Pozvánky} = 385 \div 0,18 \approx \mathbf{2\ 139\ \text{promptů}}$$

Dvě praktické implikace

01

Počítejte s dolním okrajem rozsahu.

Response raty v čase klesají, jak nastupuje únava; číslo, které fungovalo loni, bude letos nižší. Plánujte na worst case.

02

Malá populace + nízký response rate = problém.

Pokud je populace pod 1 000 a response rate pod 10 %, nedostanete obhajitelný vzorek. Buď změňte kanál (telefon, in-person, in-product), nebo přijměte, že tato skupina sama o sobě neunesse závěr 95 %/ ± 5 %.

KAPITOLA 5

Segmentace dat.

Největší past s malými vzorky v CX — jedno číslo, mnohokrát.

5

Past: jedno číslo, vynásobené.

Pravidlo 385 odpovědí platí pro vaši hlavní metriku. V okamžiku, kdy data rozkrojíte — podle regionu, segmentu, kanálu, osoby — každý řez potřebuje svých vlastních 385 odpovědí, pokud chcete stejnou přesnost pro každý. Tohle je největší past s podsamplováním v CX.

Aritmetika

Pokud chcete 95 % spolehlivost a ± 5 % chybu pro každý z K segmentů, celkový vzorek je zhruba $K \times 385$. Typický CX program, který chce krájet podle pěti zákaznických segmentů a tří kanálů — patnáct buněk — potřebuje kolem **5 775 odpovědí**, ne 385.

Tři poctivé možnosti

- 01 Akceptujte nižší přesnost pro segmentová čtení.**
 ± 10 % při 95 % vyžaduje jen 96 odpovědí na segment — v pohodě pro směrové vedení.
- 02 Predefinujte malý počet segmentů (3–5).**
 Ne všechny, co by marketing chtěl. Před spuštěním. Ne potom.
- 03 Použijte rolling 12měsíční data.**
 Pro segmenty, ne jen poslední měsíc — akumulujte víc odpovědí na buňku.

Požadavek na segment při běžných přesnostech

CHYBA ODHADU	95 % SPOL.	90 % SPOL.
± 10 procentních bodů	96	68
± 7 procentních bodů	196	138
± 5 procentních bodů	385	271
± 3 procentní body	1 068	752
± 2 procentní body	2 401	1 691

POCTIVÝ DEFAULT PRO SEGMENTY

Pro čtení na úrovni segmentu je ± 10 procentních bodů při 95 % poctivý default. Výsledek se bude hýbat 4× rychleji než vaše hlavní metrika — to je zpětná vazba, ne šum.

KAPITOLA 6

Čtení pohybů.

Kdy je změna skutečná — a kdy je to jen šum.

6

Nejmenší pohyb, který stojí za reportování.

Vaše měsíční NPS spadlo ze 41 na 37. Je to skutečné, nebo šum? Poctivá odpověď závisí na třech věcech: kolik odpovědí stojí za každým číslem, jaká je variabilita uvnitř každého období, a jestli testujete proti jedinému prahu nebo lovíte jakoukoliv změnu.

PRAKTICKÉ PRAVIDLO PRO NPS

Pro NPS je pohyb pod přibližně 5 bodů mezi dvěma měsíčními měřeními po ~400 odpovědích v rámci statistického šumu. Cokoliv pod tímto nestojí za reportováním jako „dolů“ nebo „nahoru“ bez dalšího šetření.

Jak reportovat pohyby

Fungují tři formulace. Používejte je konzistentně napříč dashboardy:

- V šumu* Když je změna pod prahem.
- Podstatný pohyb* Když překračuje práh, ale ne velkým rozdílem.
- Významný posun* Když překračuje práh dvojnásobně nebo víc.

Minimální skutečný pohyb (NPS)

Minimální skutečný posun NPS při 95 % spolehlivosti, za předpokladu stejné velikosti vzorku v obou obdobích.

ODPOVĚDÍ V KAŽDÉM OBDOBÍ	MIN. POHYB
100 každé	~14 NPS bodů
200 každé	~10 NPS bodů
400 každé	~7 NPS bodů
1 000 každé	~4 NPS body
5 000 každé	~2 NPS body
10 000 každé	~1,5 NPS bodu

JAK TO POUŽÍT

Vezměte si počet odpovědí v menším z obou období, podívejte se na řádek a porovnejte se skutečnou změnou. Pokud je váš pohyb menší než hodnota v tabulce, je to šum.

KAPITOLA 7

Šest statistických chyb.



Vzory, které vidíme v CX dashboardech nejčastěji.

Šest chyb, které CX týmy dělají se svými čísly.

Tyhle vzory vidíme nejčastěji. Žádný z nich není matematický problém — každý z nich je problém provozní disciplíny. Opravují se malým počtem pravidel, aplikovaných konzistentně napříč každým dashboardem, který tým posílá ven.

- 01** **„Máme 38 odpovědí — to stačí.“**
Nestačí. Pod ~270 odpovědích neobhájíte žádný závěr na 95 % spolehlivosti s rozumnou přesností. „Směrové čtení“ je poctivá nálepka; tak s tím zacházejte.
- 02** **Reportování pohybu bez intervalu spolehlivosti.**
Bodový odhad bez nejistoty kolem něj je defaultně zavádějící. Každá reportovaná metrika by měla cestovat se svou chybou odhadu — aspoň ve footnotě.
- 03** **Srovnávání období bez kontroly významnosti.**
Pokles ze 7,2 na 6,9 na malých vzorcích je téměř jistě šum. Porovnejte s tabulkou minimálního pohybu z kapitoly 6, než cokoliv reportujete jako trend.
- 04** **Použit celopopulační vzorek na čtení segmentů.**
Pokud je vzorek 385 odpovědí vaše jediná data, nemůžete z něj číst segmenty. Buď přijměte nižší přesnost na segment, nebo sbírejte víc odpovědí na buňku.
- 05** **Požadovat 99 % spolehlivost na všechno.**
Přepálená spolehlivost trojnásobí cenu vzorku za přesnost, kterou žádné reálné CX rozhodnutí nepotřebuje. Default 95 %; na 99 % jděte jen u výsledků, které pohnou podstatným výdajem.
- 06** **Zastavit průzkum, jakmile dosáhne magického čísla.**
Jakmile program běží, pokračujte. Průběžná data umožňují zachytit skutečný pohyb dřív; zastavení na minimálním vzorku vytváří umělá slepá místa mezi okny sběru.

KAPITOLA 8

Sebehodnocení.

Jsou vaše CX čísla skutečně obhajitelná?



Jsou vaše CX čísla skutečně obhajitelná?

Dvacet tvrzení napříč pěti oblastmi statistické disciplíny. Skórujte jeden bod za každé tvrzení, na které dnes upřímně odpovíte **ano** — ne „pracujeme na tom“.

HLAVNÍ METRIKA

- Máme aspoň 385 odpovědí za období pro hlavní metriku.
- 95% interval spolehlivosti se reportuje vedle hlavní metriky.
- Používáme stejný průzkumový nástroj napříč obdobími (žádný drift verzí).
- Umíme obhájit svou volbu velikosti vzorku před CFO.

PLÁNOVÁNÍ RESPONSE RATE

- Modelujeme počet pozvánek z požadovaných odpovědí, ne naopak.
- Předpoklad response rate bereme z dolního okraje historického rozsahu.
- Response rate sledujeme měsíčně; víme, kdy nastupuje únava.
- Když response rate klesne pod práh, průzkum pozastavíme nebo přepracujeme.

ČTENÍ SEGMENTŮ

- Každý reportovaný segment má vlastní minimální vzorek, ne vypůjčený z celku.
- Před spuštěním jsme předdefinovali malý počet segmentů — ne 20.
- Přesnost na segment reportujeme jako $\pm X$ procentních bodů, ne jen číslem.
- Pro segmenty používáme rolling okna pro akumulaci dostatečných odpovědí.

INTERPRETACE POHYBŮ

- Každá reportovaná změna se kontroluje proti prahu minimálního pohybu.
- Používáme konzistentní jazyk: „v šumu“, „podstatný“, „významný posun“.
- Mezi-obdobní srovnání obsahují test-statistiku nebo p-value, kde se to hodí.
- Jednobodové pohyby nikdy nebereme jako trend.

DISCIPLÍNA REPORTINGU

- Dashboardy ukazují intervaly spolehlivosti nebo chyby, ne jen bodové odhady.
- Velikosti vzorků jsou viditelné na každém grafu, který žene rozhodnutí.
- Máme napsaný práh, kdy pohyb spouští akci.
- Stakeholdeři vědí, co znamená „směrové“ v našich dashboardech.

JAK ČÍST SVÉ SKÓRE

16–20: Obhajitelný program; CFO a CEO se mohou spolehnout na čísla.

10–15: Hlavní metrika sedí, ale segmentová a pohybová čtení potřebují disciplínu.

Pod 10: Reportujete šum jako signál; začněte hlavním vzorkem a prahem pro pohyb.

Statistická disciplína není matematická obratnost.

Statistická disciplína v CX není o tom být chytrý se vzorci. Je o tom být poctivý k datům, která máte. Týmy, které ubrání svá čísla před skeptickým CFO, nejsou ty s nejsostikovanější matematikou — jsou to ty, které konzistentně aplikují tři pravidla.

Tři poznatky, které mají větší cenu než zbytek dokumentu:

- 01 Znej svůj minimální vzorek před spuštěním.**
Najděte řádek v tabulce z kapitoly 2, zpětně z response rate modelujte pozvánky, a teprve pak průzkum navrhnete. Ne potom.
- 02 Metrika cestuje se svou nejistotou.**
Každé číslo, které publikujete, nese chybu odhadu. Ukázat ji není příznání slabosti — je to rozdíl mezi obhajitelným reportem a tipem.
- 03 Segmentová a pohybová čtení s extra opatrností.**
Tady většina CX týmů tiše přeháněje. Pro krájení použijte tabulku na segment, pro trendy tabulku minimálního pohybu. Pokud data nepodporují tvrzení, tvrzení nedělejte.

Jak může InsightSofa pomoci.

InsightSofa aplikuje tato pravidla automaticky. Intervalvy spolehlivosti se objevují na každém grafu defaultně. Velikosti vzorků jsou viditelné všude. Čtení pohybů jsou označená proti prahu minimálního pohybu pro daný vzorek. Segmentové dashboardy odmítají zobrazit čtení pod prahem obhajitelnosti — takže váš tým nemůže omylem nadhodnotit závěry.

Napříč zhruba čtyřiceti klientskými projekty je nejčastější ranou výhrou ta samá: exekutiva, která konečně věří CX číslům, protože dashboard odmítá lhát.

Nejrychlejší start je 30minutová pracovní schůzka s naším CX strategy týmem.

+420 777 661 368 · insightsofa.com

Údaje o velikosti vzorku používají standardní vzorec pro proporce s $p = 0,5$ (worst case) a korekci konečné populace. Prahy pohybu NPS jsou přibližné, odvozené ze standardní chyby rozdílu dvou NPS skóre při stejných velikostech vzorků.