

Szemkamerás megfigyelések megbízhatósága a mintaelemszám függvényében

The reliability of eye-tracking researches depending on the sample size

Lázár Erika

PhD hallgató, Pécsi Tudományegyetem Közgazdaságtudományi Kar
lazar.erika@ktk.pte.hu

Németh Péter

Egyetemi adjunktus, Pécsi Tudományegyetem Közgazdaságtudományi Kar
nemeth.peter@ktk.pte.hu

Murai Gábor

PhD hallgató, Pécsi Tudományegyetem Közgazdaságtudományi Kar
murai.gabor@ktk.pte.hu

Szűcs Krisztián

Egyetemi docens, Pécsi Tudományegyetem Közgazdaságtudományi Kar
szucs.krisztian@ktk.pte.hu

Absztrakt

A hagyományos marketingkutatói eszköztár egyre több korlátot mutat, legyen szó akár a kérdőíves megkérdezésekről, akár a (korábban) oly népszerű fókuszcsoportos vitákról. A kiváltásuk azonban lassan megy végbe, számos objektív és néhány szubjektív kritérium miatt. Ezeket, illetve ezek közül is a megbízhatóságot emeljük ki jelen tanulmányban, amely kritikus kérdés a legtöbb kutatási projekt esetében. Ennek mértékét alapvetően a vizsgált minta nagysága alakítja, ezért kísérlettel és szimulációval vizsgáltuk meg a szemkamerás mérések működését ezen aspektusokból. Arra jutottunk, hogy a gyakorlatban elterjedt célcsoportonkénti 30-40 fős minta elfogadható megbízhatóságot eredményez, de összehasonlító elemzéseknél már a 60 fő feletti mintanagyság a javasolt.

Kulcsszavak: szemkamera, eye-tracking, mintanagyság, megbízhatóság

Abstract

The traditional marketing research toolkit has more and more limitations, even it is about questionnaires or (previously) so popular focus group discussions. However, their triggering is slow due to a number of objective and some subjective criteria. These, as well as the reliability of these, are highlighted in the present study, which is a critical issue for most research projects. The extent of this is basically shaped by the size of the examined sample, therefore we examined the operation of eye-tracking researches from these aspects by experiment and simulation. We found that 30-40 participant per target group, which is widespread in practice, results in acceptable reliability, but for comparative analyzes, a sample size of over 60 is already recommended.

Keywords: eye-tracker, eye-tracking, sample size, reliability

1. A kérdés aktualitása

Az alkalmazott marketingkutatás eszköztára folyamatosan változik, átalakul (Simon, 2016). Ennek számos oka, befolyásoló tényezője ismert (Szűcs, 2016), de jelen tanulmány szempontjából talán fontosabbak a korlátok, amelyek megnehezítik az áttörést, a hagyományos módszertani megoldásokat kiegészítő, vagy egyes esetekben akár helyettesíteni is képes technikák elterjedését.

Tudjuk, látjuk, hogy a technológia szerepe meghatározó életünk legtöbb területén, akár fogyasztóként, vásárlóként, akár gyakorló marketing szakemberként tekintünk ezekre. Működésül tartja magát a reprezentativitás kritériuma, a kvantitatív kutatások preferenciája, a "nagy számokba" vetett feltétlen bizalom. A gyakorló szakemberek többsége továbbra is akkor tartja hitelesnek a kutatási eredményeket, ha "élég sok válaszadót kérdeztünk meg".

Mindezen törekvések és tradíciók keresztmetszetében megjelentek olyan eszközök, amelyek gyors, hatékony megoldást jelentenek az adatfelvételkor, ugyanakkor egyes kritériumok mentén nem, vagy csak kevésbé felelnek meg az elvárásoknak. Jelen tanulmányban az alkalmazott marketingkutatásban már több, mint egy évtizede használt szemkamerás megfigyelések során alkalmazandó mintanagyságot vesszük górcső alá és vizsgáljuk meg a különböző minta-elemszámokhoz rendelhető megbízhatósági mutatókat.

Azt gondoljuk, hogy a szemkamerás eszköztár szakmai elfogadottságát nagyban meghatározza a kapott eredmények megbízhatósága, ebben a tapasztalataink szerint erős eltérés mutatkozik a gyakorlatban, ezért hasznos iránymutatás lehet ez az összefoglaló a vizsgálataink eredményeiről.

A szemkamerás módszertan marketingcélú felhasználására vonatkozó korábbi kutatásunk (Lázár, 2019) eredményei is azt mutatják, hogy ezen a szakterületen egyre több kutató alkalmazza az eszközt elsősorban a vásárlói magatartás és különböző kommunikációs anyagokra adott fogyasztói reakciók megértése céljából. A vizsgálódásunkból az is kiderült, hogy a tudományos gyakorlat a stabil szemkamerás módszertan használatát preferálja, vélhetően a pontosabb elemezhetőségből kifolyólag. De természetesen ez alól is vannak kivételek, a mobil szemkamerát használó kutatások jellemzően a vásárlási szituációkat vizsgálják, ahogy tették ezt Gidlő és kutatótársai (2013), amikor valós bolti környezetben vizsgálták a vásárlók döntéseit meghatározott termékkategóriákban. A tanulmány érdekessége, hogy bár manuálisan kódolták az adatokat, a pontos rögzítésnek köszönhetően statisztikailag is tudták elemezni azokat.

Általános eljárás szerint a neuromarketing módszerek esetében 30 fő körüli mintaelemszámmal dolgoznak (Szűcs - Lázár - Németh, 2019), hiszen a szakirodalomban fellelhető tanulmányok szerint ezekben az esetekben már statisztikailag megbízható eredmények születnek a ránézési adatokból (Sands, 2009). Bár részletes magyarázat korlátozottan elérhető csak a kérdésben (Lázár - Szűcs, 2020), de az ezzel foglalkozó szerzők véleménye az, hogy az agyi aktivitások alacsony szintű véletlenszerűsége miatt a résztvevők reakciói könnyebben összehasonlíthatók (Hensel et al., 2017, Genco - Pohlmann - Steidl, 2013).

Mindenesetre a tudományos kutatások területén számos példát találunk több-kevesebb elemszámmal dolgozó tanulmányra. A 30 főnél alacsonyabb mintaszámot alkalmazó tanulmányok jellemzően valamilyen kommunikációs anyag hatékonyságát mérik (Lever et al., nd., Garrison et al., 2018) weboldallal, mobilapplikációval kapcsolatos felhasználói élmény (user experience)

kérdést boncolgatnak (Cho et al., 2019, Bortko et al., 2019). Ehhez képest a magasabb elemszámmal dolgozó kutatások során egyrészt akár több célcsoport összehasonlítására is vállalkoznak, másrészt fogyasztói vagy vásárlói viselkedéseket, döntéseket írnak le (Tortora – Machin – Ares, nd, Florack – Egger – Hübner, 2020, Kwon – Adaval, 2018).

Bár az eszköz marketingkutatásban történő felhasználása - úgy tűnik - még nem tart itt, de a kapcsolódó tudományterületeken, mint a pszichológia vagy az élelmiszertudomány fogyasztói döntésekkel foglalkozó részén felmerül az egyéni jellemzők torzító hatásának kérdése, mely potenciálisan hatással lehet a mintaszám meghatározására. Gere és szerzőtársai (2017) például arra az eredményre jutottak, hogy a döntési folyamat során a ránézések idejét szignifikánsan befolyásolja az alany hangulata. Ennek következtében a pozitív hangulat hosszabb ránézéseket eredményez.

2. A kutatási kérdés és módszertani háttér

Jelen tanulmányban bemutatásra kerülő kutatásunk fókuszja, hogy az általánosan elfogadott, a legtöbb tanulmány által alkalmazott 30-40 fős mintaelemszám valóban alkalmas-e megbízható következtetések levonására a vizsgált célcsoportra vonatkozóan, vagy, ha nem, akkor mekkora mintanagyság tekinthető elegendően megbízhatónak.

A mintaelemszám teszteléséhez stabil szemkamerás vizsgálat felépítésére volt szükség, mivel jelenleg ez az eszköz ad megfelelő mennyiségű és minőségű adatot az elemzés elvégzéséhez (bár itt szükséges megjegyezni, hogy ma már a Tobii Pro mobil szemkamerás eszköze is alkalmas a felvételek statisztikai elemzésére). A terepmunka két féleven keresztül zajlott, melyben összesen 98 egyetemi hallgató vett részt (75,5%-a nő és átlagosan 21,9 évesek). Ebből kifolyólag fontosnak tartjuk kiemelni, hogy az eredmények értelmezése is erre a csoportra vonatkozóan kell, hogy megtörténjen. Mindezek ellenére érdemi megállapításokat tehetünk majd a mintaelemszám megbízhatóságát illetően, hiszen jelen esetben nem a fogyasztói reakciók elemzése volt a cél.

A kutatás laborkörülmények között zajlott, a hallgatók véletlenszerű sorrendben tekintették meg a 10-10 másodpercig vetített ingeranyagokat. Az adatfelvételhez a Tobii Pro X3-120 szemkamerát használtuk, az elemzést pedig az ehhez tartozó Tobii Pro Lab szoftver segítségével végeztük el. Az elemzésbe három kép formátumú ingeranyag került be, melyek mindegyike egy-egy ismert márka (Decathlon, MOL Fresh Corner, OTP) reklámja volt. A képek kiválasztása során fontos szempont volt a különböző típusú üzenetek és kedvezmények megjelenítése, hogy minél többféle elemet tesztelni tudjunk.

A képeket olyan részekre (Areas of Interest – AOI) osztottuk fel, mint a logó, üzenet (title/ message), konkrét ajánlat vagy kedvezmény (offer/ discount) és kiegészítő szöveges információk, melyek kisebb betűtípussal, magyarázatként szerepeltek a képeken, de megjelentek mindegyiken képes elemek (image/ face/ food) is.

Az elemzés első szakaszában Nielsen és Pernice (2009) megoldásához hasonlóan véletlen mintát vettünk 10, 20, 30, 40 és 50 főre 5-5 alkalommal. Majd ezeket az almintákat hőtérképek alapján hasonlítottuk össze.

A hőtérképek összehasonlítására építve, a második elemzési szakaszban három mutatóra nézve vizsgáltuk meg, hogyan alakul az eloszlásuk különböző mintaelemszámok esetén. Ebben az esetben 10.000 véletlen kiválasztást alkalmaztunk 15, 30, 45, 60 és 75 fő esetében minden AOI tekintetében az alábbi mutatókat érintve:

- teljes ránézési idő (Total Fixation Duration),
- átlagos ránézési idő (Average Fixation Duration),
- az első ránézésig eltelt idő (Time to First Fixation).

1. ábra: A vizsgált ingeranyagok



Forrás: saját szerkesztés

3. Kutatási eredmények

Az elemzés első szakaszának elvégzésének célja annak megállapítása volt, hogy a Nielsen és Petrice (2009) által írt tanulmányban manuálisan, hőterképekkel végzett mintaelemszám összehasonlítás mennyiben állja meg a helyét a saját kutatásunkban. A szerzőkhöz hasonlóan mi is 10, 20, 30, 40 és 50 fős mintákat vettünk véletlenszerűen, de minden esetben 5 alkalommal. A képek most a Decathlon ingeranyagát szemléltetik, de az eredmények minden reklámanyag esetében hasonló képet mutattak.

A hőterképeket a teljes vetítési időre nézve (10 mp) kezeltük és a ránézések időtartamára vonatkozóan (absolute duration) ábrázoltuk.

2. ábra: Decathlon ingeranyag hőterképes vizualizációja 10 másodpercre vetítve (n=98)



Forrás: saját szerkesztés

A különböző elemszámmal vett hőterképek (3-7. ábra) alapján látszik, hogy a hivatkozott Nielsen és Petrice (2009) tanulmányhoz hasonlóan a mintavétel valóban úgy tűnik, hogy a minták közötti különbség 30-40 fő felett kezd el szemmel láthatóan csökkenni. 10 és 20 fő esetében az 5 mintavétel között még jelentős eltérések vannak, míg 30, 40 és 50 fő tekintetében az 5-5 minta között kisebb az eltérés és a 98 fős hőterképhez hasonló következtetéseket tudnánk levonni belőlük. Ez azonban ebben a formában egy kvalitatív megfigyelés inkább, mintsem egy erős állítás, hiszen mindössze 5-5 mintavételt végeztünk el.

3. ábra: Decathlon ingeranyag hőterképes vizualizációja 10 másodpercre vetítve (n=10)



Forrás: saját szerkesztés

4. ábra: Decathlon ingeranyag hőterképes vizualizációja 10 másodpercre vetítve (n=20)



Forrás: saját szerkesztés

5. ábra: Decathlon ingeranyag hő térképes vizualizációja 10 másodpercre vetítve (n=30)



Forrás: saját szerkesztés

6. ábra: Decathlon ingeranyag hő térképes vizualizációja 10 másodpercre vetítve (n=40)



Forrás: saját szerkesztés

7. ábra: Decathlon ingeranyag hőterképes vizualizációja 10 másodpercre vetítve (n=50)



Forrás: saját szerkesztés

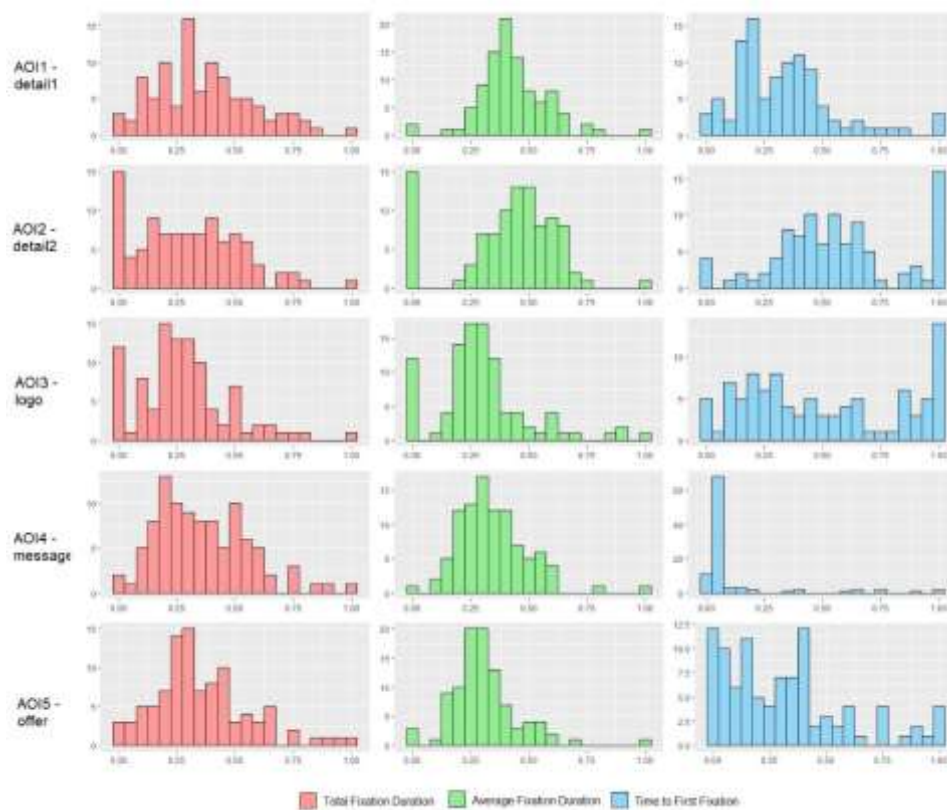
Látszik tehát, hogy vannak különbségek attól függően, hogy milyen mintával dolgozunk a szemkamerás kutatás során, de ahhoz, hogy meghatározzuk, hogy milyen valószínűséggel követünk el hibát a minta kiválasztása során, szimulációra volt szükség, melyben ingeranyagként 10.000-szer vettünk mintát.

Ahhoz azonban, hogy közelebb kerüljünk a megfelelő mintanagyság megállapításához, a mintabeli eloszlások összehasonlítását végeztük el. A központi határeloszlás tétel értelmében a mintából számolt átlag várható értéke megegyezik a populációs várható értékkel, varianciája pedig egyenlő a populáció varianciájának és a mintaelemszám hányadosával (WOOLDRIDGE, 2016). Ez azt jelenti, hogy a mintaelemszám növelésével csökken a mintából számolt átlag várható eltérése a populációs átlagtól. Az eltérés csökkenésének mértéke minden további elem hozzáadásával exponenciálisan csökken, tehát minden újabb elem kevésbé lesz hasznos. A mintából számolt átlag megbízhatósága a populációs variancia és a mintaelemszám függvénye.

A mintavételi átlag eloszlások elemenkénti és mutatónkénti összehasonlításához az adatok normalizálására volt szükség (8-10. ábra), ahol a 0 a legkisebb 1 pedig a legnagyobb érték. A hisztogramok azt mutatják meg, hogy miként oszlanak meg a ránézések a különböző AOI-k esetében. A teljes és az átlagos ránézések időtartama esetében a 0 érték a rá nem nézést jelentik és az érték minél közelebb van az 1-hez annál nagyobb a ránézési érték, így egyértelműen látszanak azok az AOI-k, melyeket jobban figyelembe vettek a résztvevők.

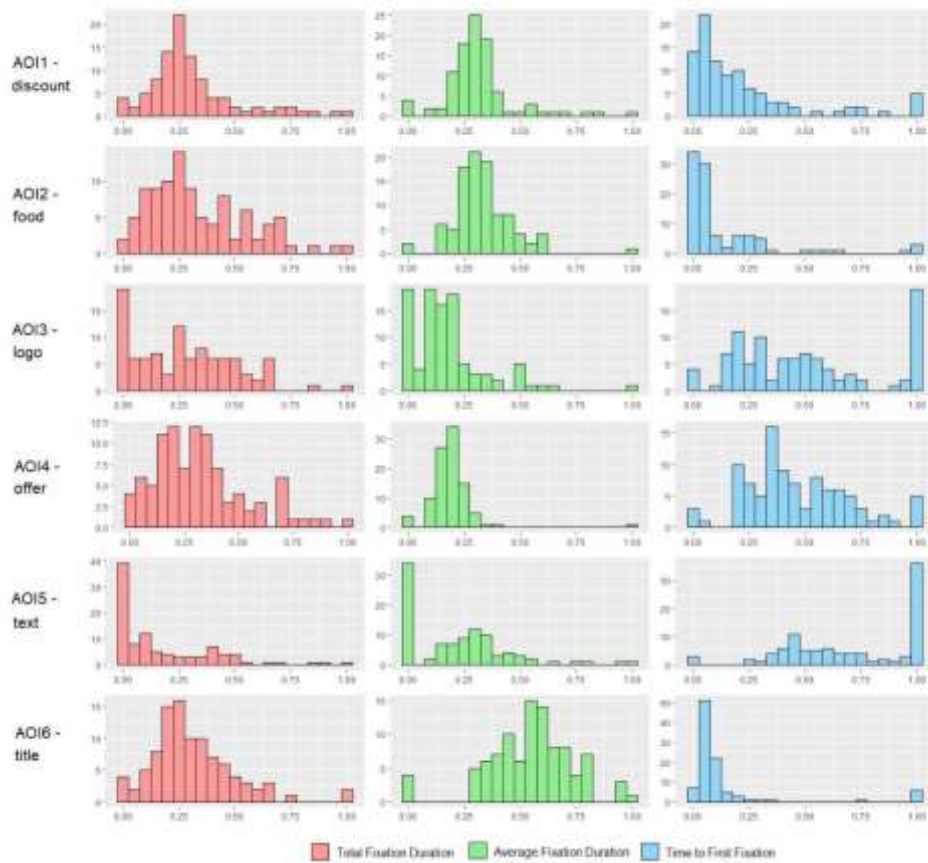
Ebben az esetben az első ránézésig eltelt idő fordítva működik, hiszen minél kisebb az érték, annál kevesebb idő telt el az első megtekintésig és itt az 1-es érték jelenti a rá nem nézést. Ennek következtében lesz majd ennek a mutatónak az eloszlása folyamatosan laposabb.

8. ábra: Normalizált mérési eredmények hisztogramja: a Decathlon ingeranyag esetében (n=98)



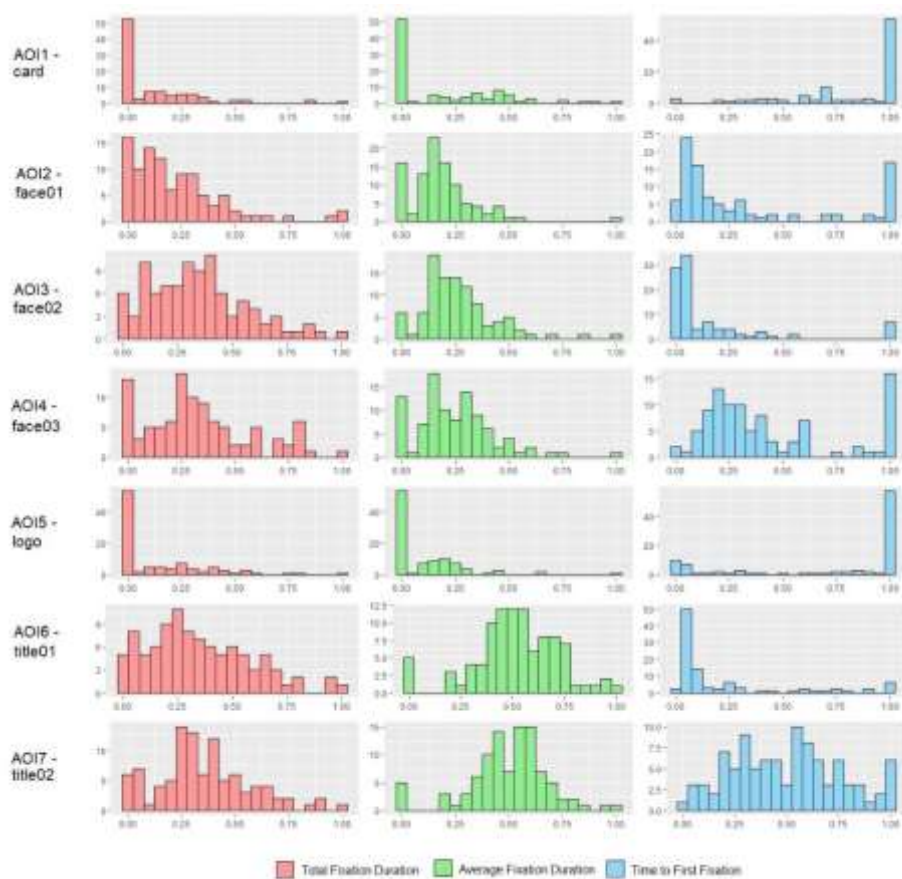
Forrás: saját szerkesztés

9. ábra: Mintavételi átlag eloszlása a MOL Fresh Corner ingeranyag esetében (n=98)



Forrás: saját szerkesztés

10. ábra: Mintavételi átlag eloszlása az OTP ingeranyag esetében (n=98)



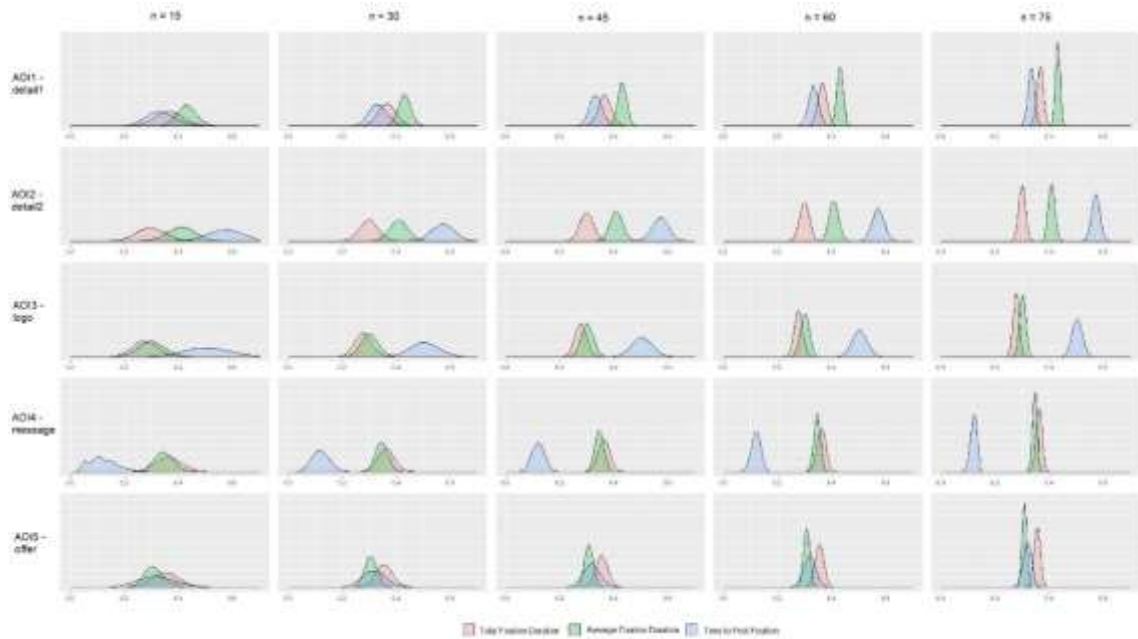
Forrás: saját szerkesztés

Az eredeti eloszlás figyelembevételéért fontos, mert ez nagyban befolyásolja a mintabeli eloszlást is. Az eredmények alapján egyértelműen látszik, hogy a 30 fő feletti mintaelemszámok esetében jelentős javulást mutatnak az adatok, így 60 fő felett egészen megbízhatóak az eredményeink (11-13. ábra).

Azonban az is egyértelműen látszik, hogy az adatok megbízhatósága nagyban függ:

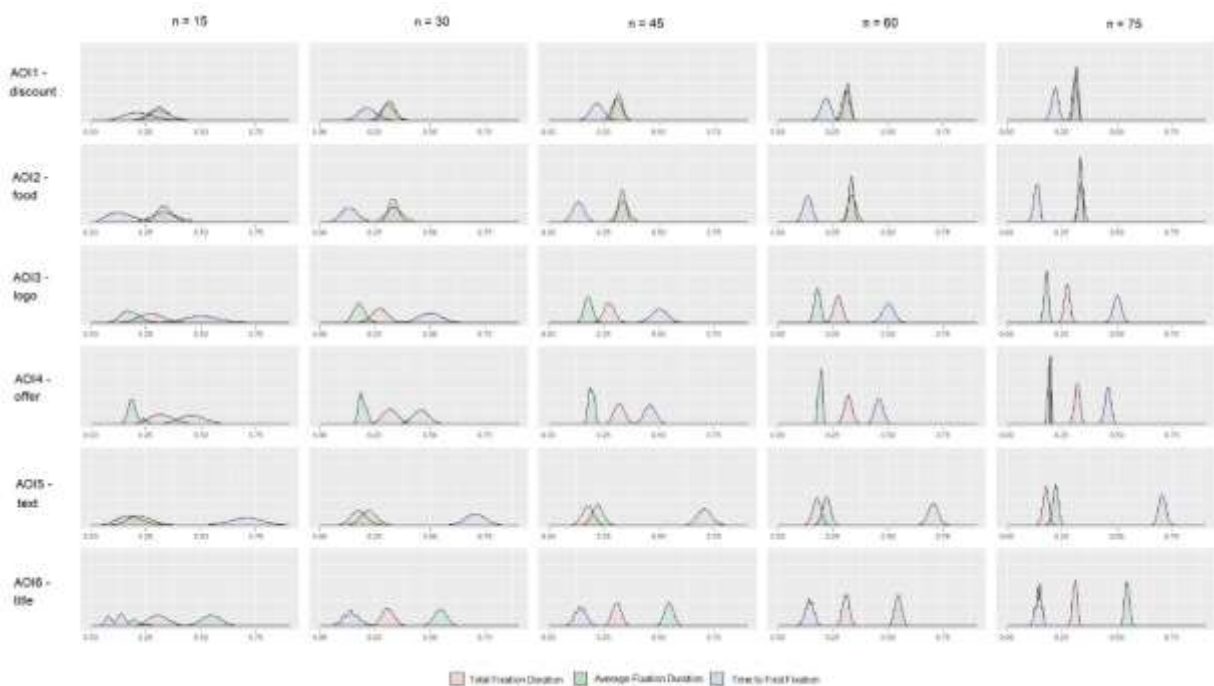
- az ingeranyag típusától: minél aprólékosabban jelöljük ki az AOI-kat annál kevésbé lesz pontos egy 30 fős minta, míg a jól elhatárolható, nagyobb elemekkel rendelkező ingeranyagoknál az eloszlás jóval tisztább képet mutat,
- az AOI típusától: számokat, nagyobb betűméretet és képeket pontosabban tudunk mérni 30 fővel, míg kisebb vagy hosszabb üzenetek mérésére érdemes nagyobb elemszámon gondolkodni,
- és a választott mutatótól is: az eredmények alapján az első ránézésig eltelt idő (Time to First Fixation) a leginkább érzékeny, míg az átlagos ránézési idő (Average Fixation Duration) 30-45 fő esetében a legpontosabb.

11. ábra: Mintabeli átlag eloszlása mintaelemszámonként és AOI-nként: Decathlon ingeranyag



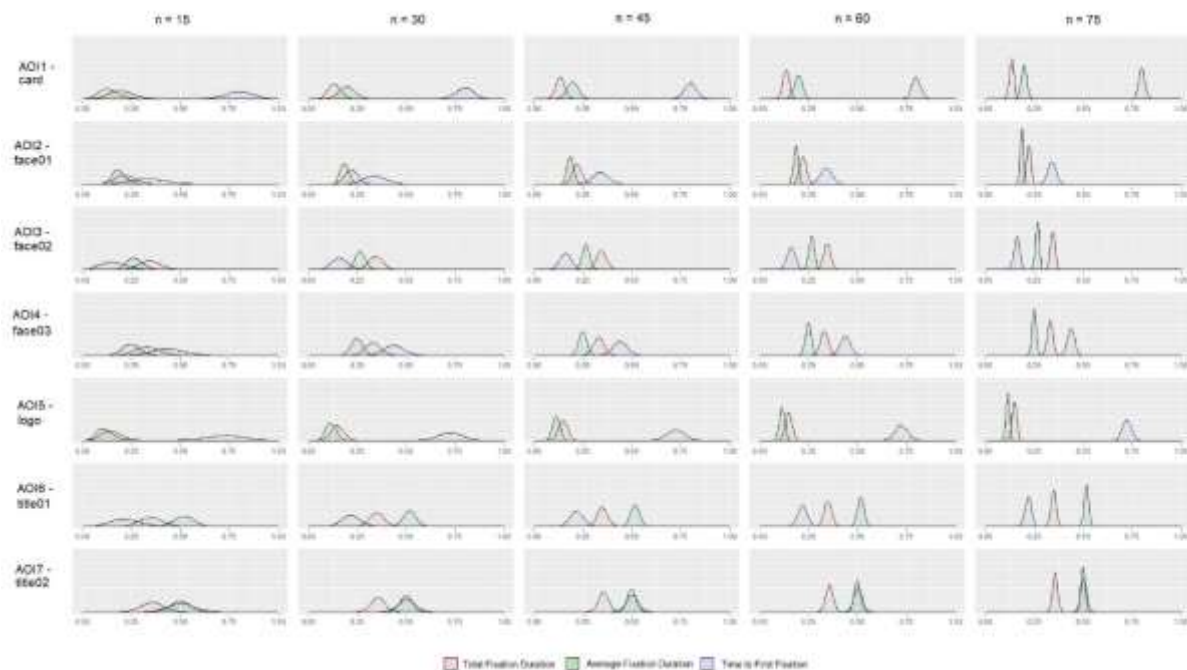
Forrás: saját szerkesztés

12. ábra: Eloszlások mintaelemszámonként és AOI-nként: MOL Fresh Corner ingeranyag



Forrás: saját szerkesztés

13. ábra: Eloszlások mintaelemszámoként és AOI-nként: OTP ingeranyag



Forrás: saját szerkesztés

4. Konklúziók

Jelen tanulmány a mintaelemszámmal kapcsolatos empirikus vizsgálódásunk első állomása, mely számos korlátra és további kérdésre világított rá.

Az adatfelvételt egyetemi célcsoport körében rögzítettük, így a megállapításainkat is ennek megfelelően kell kezelnünk. Fontos lesz a jövőben annak felmérése, hogy az ezen a mintán kapott eredmények mennyiben általánosíthatók más korosztályokra vagy célcsoportokra. Mondhatjuk-e vajon azt, hogy minden célcsoportban elegendő egy bizonyos mintaelemszám vizsgálata?

Az elsődleges elemzéseket statikus, kép formátumú ingeranyagokon végeztük el, de a vizsgálat tartalmazott videóanyagot is. Az már most kiderült, hogy az AOI tartalma és típusa meghatározza a ránézés megbízhatóságát, így ez várhatóan a dinamikus AOI-k esetében is érdekes felvetés lesz.

Az eredeti tervekkel (n=120) ellentétben a tanulmány elkészültéig 98 fő mérése valósult meg, ezt a mintát tekintettük populációnak, ami magával vonja a kérdést, hogy ez a minta mennyire tekinthető pontosnak. Így egyértelmű cél a jövőben az alapminta bővítése.

Mindezen korlátok és jövőbeli kutatási irányok mellett természetesen vannak olyan megállapítások, amelyek jelen kutatás eredményeiként fogalmazhatók meg. Megerősítést nyert, hogy az alkalmazott kutatások esetében gyakori 30 fő körüli mintanagyság elfogadható megbízhatóságot jelent, főként, ha az elemzések egy célcsoportra és az általánosan használt mutatókra koncentrálnak. Fontos eredmény, hogy a megbízhatóság szoros korrelációt mutat a kutatás fókuszával, azaz, minél pontosabban tudunk mérési területet kijelölni, annál megbízhatóbb eredményeket fogunk kapni. A statikus ingeranyagok mérése – az előző feltételek teljesülése esetén – nagyobb megbízhatósággal végezhető el kisebb mintán ugyanazon célcsoport eseté-

ben. Ebből következik az is, hogy lehetőség szerint kerülni kell több célcsoport együttes vizsgálatát alacsony elemszámú mintavétel esetén, mert jelentős torzítások jelenhetnek meg, amelyek az eredmények félreértelmezéséhez vezethetnek.

Irodalomjegyzék

- Bortko, K., Bartkó, P., Jankowski, J., Kuras, D. & Sulikowski, P. (2019). Multi-criteria Evaluation of Recommending Interfaces towards Habituation Reduction and Limited Negative Impact on User Experience, *Procedia Computer Science*, 159, pp. 2240–2248. doi: 10.1016/j.procs.2019.09.399.
- Cho, H., Powell, D., Pichon, A., Kuhns, L. M., Garofalo, R. & Schnall, R. (2019): Eye-tracking. g Retrospective Think-aloud as a Novel Approach for a Usability Evaluation. doi: 10.1016/j.ijmedinf.2019.07.010
- Florack, A., Egger, M. and Hübner, R. (2020). When products compete for consumers attention: How selective attention affects preferences, *Journal of Business Research*, 111, pp. 117–127. doi: 10.1016/j.jbusres.2019.05.009.
- Garrison, K. A., O'Malley, S. S., Gueorguieva, R. & Krishnan-Sarin, S. (2018). A fMRI study on the impact of advertising for flavored e-cigarettes on susceptible young adults, *Drug and Alcohol Dependence*, 186, pp. 233–241. doi: 10.1016/j.drugalcdep.2018.01.026.
- Gere, A., Kokai, Z. and Sipos, L. (2017). Influence of mood on gazing behavior: Preliminary evidences from an eye-tracking study', *FOOD QUALITY AND PREFERENCE*, 61, pp. 1–5. doi: 10.1016/j.foodqual.2017.05.004.
- Genco, S. J., Pohlmann, P. A. & Steidl, P. (2013). Neuromarketing For Dummies
- Gidlöf, K., Wallin, A., Dewhurst, R. & Holmqvist, K. (2013). Using Eye Tracking to Trace a Cognitive Process: Gaze Behaviour During Decision Making in a Natural Environment, *Journal of Eye Movement*, 6(1)
- Kwon, M. and Adaval, R. (2018). Going against the Flow: The Effects of Dynamic Sensorimotor Experiences on Consumer Choice, *Journal of Consumer Research*, 44(6), pp. 1358–1378. doi: 10.1093/jcr/ucx107.
- Lever, M. W., Shen, Y. and Joppe, M. (no date). Reading travel guidebooks: Readership typologies using eye-tracking technology', *JOURNAL OF DESTINATION MARKETING & MANAGEMENT*, 14. doi: 10.1016/j.jdmm.2019.100368
- Lázár E. & Szűcs K. (2020). A neuromarketing aktuális helyzete és a mintaelemszámra vonatkozó kihívásai, különös tekintettel a szemkamerás mérésekre, *Vezetéstudomány / Budapest Management Review*, 51(3): 79-88. p
- Nielsen, J., & Pernice, K. (2009). How to conduct eye tracking studies. Letöltve [https://media.nngroup.com/media/reports/free/How to Conduct Eyetracking Studies.pdf](https://media.nngroup.com/media/reports/free/How_to_Conduct_Eyetracking_Studies.pdf)
- Sands, S. F. (2009): Sample Size Analysis for Brainwave Collection (EEG) Methodologies. Letöltve: <http://www.sandsresearch.com/assets/white-paper.pdf>
- Hensel, D., Iorga, A., Wolter, L. & Znanewitz, J. (2017): Conducting neuromarketing studies ethically practitioner perspectives. *Cogent Psychology*, 4(1): 1-13
- Szűcs, K. (2016): Marketingkutató 2.0. *Vezetéstudomány-Budapest Management Review*, 47(4), 67-70.
- Szűcs K., Lázár E. & Németh P. (2019). The market position of eye tracking and its challenges in sample size. In: János, Steklács; Zsóka, Sipos; Szilvia, T. Varga (szerk.) Abstracts of Presentations : 4th Hungarian Eye-Tracking Conference. 31th May 2019, Budapest.

Tortora, G., Machin, L. and Ares, G. (no date). Influence of Nutritional Warnings and Other Label Features on Consumers' Choice: Results from an Eye-Tracking Study. *FOOD RESEARCH INTERNATIONAL* 119: 605–11. doi:10.1016/j.foodres.2018.10.038.

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach*. Nelson Education.

Zhang, X., Fan, C., Yuan, S. & Peng, Z. (2015). An Advertisement Video Analysis System Based on Eye-Tracking