

OTDK-dolgozat

2025

BUDAPESTI CORVINUS EGYETEM



**MENNI VAGY NEM MENNI? – A BALATONI ÉS ADRIAI
SZÁLLÁSÁRAK ÖSSZEHAJONLÍTÁSA PROPENSITY SCORE
MATCHINGGEL**

**TO GO OR NOT TO GO? – COMPARING ACCOMMODATION
PRICES BETWEEN LAKE BALATON AND THE ADRIATIC USING
PROPENSITY SCORE MATCHING**

Dócs Dávid Zoltán, Hajdú Marcell

Konzulens: Madari Zoltán

Kézirat lezárásának dátuma: 2025. 01. 15.

Absztrakt

Kutatásunkban arra a kérdésre kerestük a választ, hogy vajon még mindig több pénzbe kerülnek-e a szállások a horvát Adriai-tenger mentén, mint a Balatonon. A kérdés megválaszolására a Booking.com-ról web scraping segítségével letöltött adatokon dolgoztunk és propensity score matching módszert alkalmaztunk az árkülönbség ellenőrzésére. Összesen 847 szállást vizsgáltunk. Az adataink 2 felnőtt részére, 6 éjszakára szóltak és 2025 júliusára vonatkoztak. Lineáris regresszió segítségével meghatároztuk, hogy mely változók állnak szignifikáns kapcsolatban a szállásárakkal. Az így kiválasztott változókkal logit modell alapján hozzárendeltünk minden szálláshoz egy propensity score-t, majd nearest neighbor matching módszerrel összepárosítottuk az egymáshoz leginkább hasonló jellegű balatoni és horvát szállásokat. A matchinget követően sikerült kiegyensúlyozni a két hely szállásai között a változók eloszlását, így kontrolláltunk az árkülönbség egy részét okozó összetétel hatásra. Az összepárosított szállások közötti árkülönbséget Welch-próbával teszteltük. A teszt eredményeként azt kaptuk, hogy a szállaspárok közötti átlagos árkülönbség nem szignifikáns. Ebből azt a konklúziót tudtuk levonni, hogy a szálláshelyek közti alapvető árkülönbségeket nem az eltérő ország, hanem a különböző szállásjellemzők eredményezik, mint például a konyha megléte, a wellness szolgáltatások, vagy a vízparthoz közelebbi elhelyezkedés. Az előzetes hipotézisünket elutasítottuk, miszerint a horvát szállások drágábbak, mint a balatoni szállásárak. Az országok közötti árkülönbségek mögött leginkább az állhat, hogy Horvátországban több magas felszereltséggel, illetve vízparthoz közelebbi elhelyezkedéssel rendelkező szálláshely található.

Kulcsszavak: szállásárak, propensity score matching, Balaton, Horvátország, Adria

JEL klasszifikáció: C14, C21, C55, D12, L83

Tartalomjegyzék

Absztrakt.....	i
Ábrák- és táblázatok jegyzéke.....	iii
1. Bevezetés.....	1
2. Szakirodalmi összefoglaló	2
2.1. Turisztikai imázs	2
2.2. Balaton, mint turisztikai célpont	3
2.3. Horvátország, mint turisztikai célpont	3
2.4. Szállásárakra ható tényezők.....	4
3. Adatok	6
4. Módszertan.....	11
4.1. Többváltozós lineáris regresszió	11
4.2. Propensity score matching	12
4.2.1. Logisztikus regresszió.....	12
4.2.2. Matching	13
4.3. Welch-próba	14
5. Eredmények	14
5.1. Lineáris regressziós modell	15
5.2. Propensity score kiszámítása logit modellel.....	18
5.3. Matching.....	19
5.4. Kétmintás várható érték próba	21
6. Kutatás limitációi	22
7. Összefoglalás	23
8. Irodalomjegyzék	25

Ábrák- és táblázatok jegyzéke

Ábrák

1. ábra: Távolság mérése a vízparttól.....	7
2. ábra: Balatoni szálláshelyek az adatbázisban.....	8
3. ábra: Horvát szálláshelyek az adatbázisban.....	9
4. ábra: A szállásárak hisztogramja.....	10
5. ábra: A szállásárak logaritmusának hisztogramja.....	15
6. ábra: A propensity score-ok eloszlása országonként.....	19
7. ábra: A propensity score-ok eloszlása a teljes és a párosított mintában, országonként.....	20
8. ábra: A standardizált átlagkülönbségek.....	21

Táblázatok

1. táblázat: Változólista.....	6
2. táblázat: Folytonos változók leíró statisztikái.....	9
3. táblázat: A dummy változók megoszlása országonként.....	10
4. táblázat: Lineáris regressziós modellek.....	15
5. táblázat: Logit modell.....	18

1. Bevezetés

Számos magyar család és egyén számára a nyári időszak központi eleme a vízparti vakáció. A magyarok számára a két legvonzóbb célpont évek óta a Balaton térsége, valamint a horvát Adriai-tenger partszakasza menti nyaralás. 2024-ben a nyári hónapokban a balatoni turisztikai térségben volt a szálláshelyeknek a legnagyobb forgalma Magyarországon, ahol a nyár folyamán több, mint 3,9 millió vendégéjszakát töltöttek el pihenéssel a magyarok (KSH, 2024a). A külföldi, tengerparttal rendelkező országok közül Horvátország volt az első számú célpont, ahova 2023 nyarán közel 630 ezer utazást tettek a magyarok (KSH, 2024b).

A Covid-járványt követően újból beinduló balatoni turizmusra az évről-évre történő jelentős áremelkedés volt jellemző. A Pénzcentrum (2024a) elemzése szerint 2023-ról 2024-re a szállásárak közel 10%-kal emelkedtek a Balatonon az előző év azonos időszakához képest. A növekvő tendenciájú balatoni árszínvonal következtében egyre nagyobb figyelmet kapott a horvátországi nyaralás, mint alternatíva a balatoni üdüléssel szemben.

A szembetűnő áremelkedés kapcsán az idei évben számos internetes cikk (24.hu, 2024; RTL.hu, 2024; Pénzcentrum, 2024b) és videós riport (Telex.hu, 2024) készült annak a kérdésnek a megválaszolására, hogy a balatoni vagy a horvátországi nyaralásnak van-e alacsonyabb anyagi vonzata. Ezek a munkák többnyire spekulatív jellegűek, és/vagy a teljesség igénye nélkül csupán egy-egy szálláshely költségeit hasonlítják össze. Kevés tudományos jellegű munka jelent meg a témában. Az egyik ilyen kutatás - amelyre az említett cikkek is előszeretettel hivatkoznak - a Balaton-Adria nyaralás költségeinek összehasonlítását nagyobb mintaelemszám figyelembevételével elemzi, és dr. Sulyok Judithoz, a Pannon Egyetem Turizmus Intézeti Tanszékének tudományos főmunkatársához köthető. A kutatás publikáció formájában még nem jelent meg, az eredményeket egyedül a Portfolio.hu által rendezett Balaton Konferencia 2024-en egy prezentáció keretein belül mutatta be a szerző (Portfolio.hu, 2024).

A fentiekből kifolyólag a kutatásunk mögötti motivációnk az, hogy a magyarok nyári üdülési célpont választás kérdésében létrehozzunk egy olyan releváns tudományos munkát, amely fejlett statisztikai módszerekkel és nagy mintaelemszám segítségével széleskörű összehasonlítást tesz lehetővé a balatoni és a horvátországi szálláshelyek árai között. Kutatási célunk, hogy átfogóan elemezzük és vizsgáljuk meg a balatoni és horvát partmenti szálláshelyek közötti árkülönbségeket és a különbségek mögöttes magyarázó tényezőit a 2025-

ős év nyarára foglalható Booking.com-on található szállásadatok alapján. További célunk, hogy megpróbáljunk választ nyújtani arra az előzetes hipotézisre és feltevésre, miszerint a horvát Adriai-tenger menti szállásárak drágábbak, mint a balatoni szálláshelyek. Kutatásunk a szállások árára fókuszál, az utazással kapcsolatos egyéb költségek nem képezik elemzésünk részét.

Dolgozatunkban először a releváns szakirodalmi cikkeket szeretnénk bemutatni, majd a 3. fejezetben a felhasznált adatokat ismertetjük, részletezve az adatgyűjtést és az adatok leíró statisztikáit, térinformatikai eszközökre is támaszkodva. Ezt követően a 4. fejezetben a használt módszertant mutatjuk be, majd az 5. fejezetben az eredmények mélyebb értelmezését végezzük el. A hatodik fejezetben megvitatjuk kutatásunk limitációit és továbbfejlesztési lehetőségeit, majd az utolsó fejezetben összefoglaljuk a kapott eredményeinket és következtetéseinket.

2. Szakirodalmi összefoglaló

A szakirodalmi összefoglalót négy fő téma köré csoportosítottuk. Először összefoglaljuk turisztikai szemszögből egy célpont imázsának az értelmezését. Aztán áttekintjük a Balaton térségét, mint turisztikai célpontot. Majd szemügyre vesszük a horvát Adria-tengerpartot, mint üdülési helyszínt. Végül megvizsgáljuk, milyen jellemzők állhatnak kapcsolatban egy adott szálláshely árával.

2.1. Turisztikai imázs

Gerdesics (2017) szerint az imázs egy olyan kép, amely egy adott vállalatról, termékről vagy márkáról az emberekben elképzelések és benyomások összességeként kialakul. A turisztikai imázs szerepe az utazási desztinációk megítélésében és népszerűségében kiemelkedő jelentőséggel bír. A turisztikai imázs többdimenziós jelenség, amely az adott hely természeti adottságaitól kezdve a kulturális örökségen át a helyi lakosok vendégszeretetéig számos tényezőt ölel fel. Az imázs alapját a kognitív és az affektív elemek adják, vagyis egyrészt a desztinációval kapcsolatos információk és ismeretek, másrészt pedig az ehhez kapcsolódó érzelmi reakciók. Ezek az elemek hatással vannak arra, hogy a potenciális látogatók mennyire érzik vonzónak az adott helyet, és hogy milyen elköteleződést alakítanak ki vele szemben (Wu, 2016; Moreno-González et al., 2020).

Sulyok és Papp (2023) rávilágítanak arra, hogy a turisztikai imázs alakításában nemcsak a közvetlen élmények játszanak szerepet, hanem a közvetett információforrások is, mint például a közösségi média, az online értékelések, valamint a reklámok és médiatartalmak. Ezek a

források hozzájárulnak ahhoz, hogy egy desztinációról komplex kép alakuljon ki, amely meghatározza a turisták elvárásait és elégedettségét (Sulyok és Papp, 2023). A közösségi média egyre fontosabbá válik a fiatalabb generációk számára, akik előszeretettel építenek mások tapasztalataira (Dedeoğlu et al., 2020).

A turisztikai imázs azonban nem mindig a valóságot tükrözi, mivel gyakran sztereotípiák és kulturális előítéletek is hatnak rá. Sulyok és Papp (2023) is alátámasztják, hogy a turisztikai imázs megváltoztatása hosszú távú erőfeszítéseket igényel, és általában célzott kommunikációs stratégiák révén érhető el, amelyek a helyi kulturális és természeti adottságokra építenek, valamint az egyedi élményeket hangsúlyozzák.

2.2. Balaton, mint turisztikai célpont

2023-ban a balatoni turisztikai térségben közel 6,3 millió vendégéjszakát töltöttek a magyar emberek: ezzel a legnépszerűbb turisztikai térségnek számított a Balaton környéke a belföldi vendégéjszakák alapján (KSH, 2024a). Ez nem feltétlen meglepetés, tekintve, hogy a Balaton Közép-Európa legnagyobb tava, valamint Magyarország legjobban kiépült turisztikai térsége (Domonkos et al., 2016).

Domonkos és szerzőtársai (2016) sokrétűen elemzik a Balaton idegenforgalmi régiót elsősorban gazdasági szempontból. Azt találták, hogy a térség turizmusa jelentős gazdasági és társadalmi hatással bír Magyarországon, mivel a szálláshelyek és vendéglátóipari szolgáltatások mellett a fürdők és szabadidős tevékenységek is hozzájárulnak a régió bevételeihez. A balatoni turizmus nemcsak közvetlen bevételi forrást biztosít, hanem közvetetten is támogatja a helyi gazdaságot, például munkahelyteremtés és infrastruktúra-fejlesztés révén. A Balaton sajátos természeti és kulturális adottságai, mint a borászat és a wellness, erősítik a térség versenyképességét és a fenntartható fejlődés lehetőségét.

Sulyok és Papp (2023) a Balatont a turisztikai imázs perspektívájából vizsgálják meg. Sulyok és Papp (2023) szerint a térség legfőbb vonzereje a természeti szépség és a biztonság, amelyek a hazai látogatók számára különösen pozitív imázselemek. Ugyanakkor a Balaton drágasága és zsúfoltsága problémaként jelenik meg mind a látogatók, mind a helyi szakemberek szemében. A turizmushoz kapcsolódó fejlesztések és a közösségi média szerepe hozzájárulnak a desztináció pozitív megítéléséhez és fontos eszközei az imázs alakításának.

2.3. Horvátország, mint turisztikai célpont

2023 nyarán Horvátország volt az elsőszámú tengerparttal rendelkező külföldi célpont, ahova közel 630 ezer utazást tettek a magyarok (KSH, 2024b). A KSH (2024b) adatai szerint a

nyár folyamán a magyarok összesen több, mint a 2,9 millió napot töltöttek itt el, és több, mint 60 milliárd Ft-ot költöttek el Horvátországba tett utazásuk során. Rácz (2016) szerint Horvátországon belül az eltöltött vendégéjszakák 95%-a tartósan az adriai térségre koncentrálódik. Mindezek ellenére nincsen tág szakirodalma a horvátországi nyaralásoknak, se a horvát turisztikára irányuló tudományos munkáknak.

Gerdesics (2017) is értekezésében az országimázs irányából közelítette meg Horvátországot. Munkájában hangsúlyt fektet arra, hogy magyar szemmel milyen országimázssal rendelkezik Horvátország és kérdőíves mintavételezéssel elemzi a Horvátországgal kapcsolatos asszociációkat. A Horvátországról alkotott magyarországi asszociációk leginkább a turizmussal és a tengerparttal kapcsolatosak, ami tükrözi az ország domináns turisztikai imázsát. A 681 válaszadó 83,1%-a a „tenger” „Adria” és más tengerparti nyaraláshoz köthető kifejezéseket említette, amelyek semleges tartalmúnak tekinthetők. Negatív asszociációk, mint a „háború” vagy a „Balkán,” csak 1,6%-ban jelentek meg, míg a sport, amelyet a szakirodalom fontos imázsalkotó elemként tart számon, csupán az esetek 1%-ában szerepelt. Megállapítható, hogy összességében a magyarok számára Horvátország főként egy tengerparti, nyári turisztikai célpontot jelent.

2.4. Szállásárakra ható tényezők

A különböző szálláshelyek árképzésének megértésére számos elmélet született. A hotelek árazásának 4 fő típusát különbözteti meg Hung et. al (2010): költség alapú árazás, verseny alapú árazás, fogyasztó alapú árazás és érték alapú árazás. Ezek a megközelítések elsősorban kínálati alapú elméleti összefüggések, amelyek azt tárgyalják, hogy a szállás tulajdonosa milyen stratégiák mentén tudja meghatározni a hotelszobáinak árait.

A mi kutatásunk szempontjából relevánsabbak az úgynevezett hedonikus árazási modellek. A hedonikus árazás elméleti hátterét Rosen (1974) alkotta meg, aki szerint egy adott termék árát a termék jellemzőinek együttese határozza meg. Ahhoz, hogy az egyes jellemzők árhoz való hozzájárulásuk mértékét vizsgálni tudják, leggyakrabban lineáris regresszió alapuló becsléseket alkalmaztak, melyekben a termék árát a termék tulajdonságainak lineáris függvényeként modellezték.

Számos kutatás irányult a különböző szálláshelyek (hotelek, apartmanok, lakások) árai és jellemzői közötti összefüggések keresésére a hedonikus árazási modellek segítségével. Hartman (1989) például azt találta, hogy a szobaszervíz szolgáltatás és az ingyenes parkolási lehetőség várhatóan emeli a hotelszobák árát. Thrane (2007) osloei adatokon szintén megállapította, hogy a hotelszoba ára és a parkolási lehetőség megléte között pozitív irányú

összefüggés állhat fent. Továbbá a kutatásából kiderül, hogyha egy hotel minél közelebb van a város központjához, valamint a hotelszoba rendelkezik minibárral vagy hajszáritóval, jellemzően annál több pénzt kell költenie a szolgáltatást igénybe vevő egyéneknek. Israeli (2002) és Bull (1994) mindketten arra az eredményre jutottak, hogy a hotel csillagainak száma pozitív irányú összefüggést mutat a szobák árával a többi tényező változatlansága mellett. Bull (1994) továbbá ausztrál adatok elemzése alapján szintén azt találta, hogy a városközponttól vett távolság szignifikáns negatív irányú kapcsolatban áll az árral.

Az általunk vizsgált két célterületen kiemelten kevés szállásárazással kapcsolatos tanulmány készült. A Balaton környéki szállásokra Dudás és szerzőtársai (2020) munkája irányult, míg a horvát térségre nem találtunk ilyen témájú releváns kutatást. Dudás és szerzőtársai (2020) Airbnb szállások árait futtattak lineáris és kvantilis regressziókat, ahol magyarázó változóként az adott szállások jellemzői szolgáltak. Arra az eredményre jutottak, hogy a légkondicionálás, az ingyenes vezeték nélküli internet, az ingyenes parkolás, valamint a hálósobák és fürdőszobák száma szignifikánsan meghatározza a balatoni Airbnb szállások árának várható értékét.

Ahogy korábban is említettük, a balatoni és horvát szállásárak összehasonlítására dr. Sulyok Judit kutatását találtuk relevánsnak, amelynek eredményei kizárólag a Portfolio.hu által szervezett Balaton Konferencia 2024 esemény prezentációjában kerültek bemutatásra (Portfolio.hu, 2024). A Portfolio.hu (2024) cikkének tudósítása alapján Sulyok 7 éjszakára vonatkozóan vizsgálta meg a szállásokat. A kutatásban Sulyok csak a legnépszerűbb balatoni városokat vetette össze a minimum 40 ezer vendégéjszakát realizáló horvát településekkel. Sulyok arra az eredményre jutott, hogy egyedül családi nyaralás (két felnőtt, két gyerek) esetében olcsóbb Horvátország, feltéve, hogy a legolcsóbb szállaskategóriát választják a nyaralók, amely a kempingszállás önellátással. A drágább árkategóriákban viszont azt találta Sulyok, hogy költségesebb Horvátország: egy négytagú családnak 600 ezer és 1,8 millió forint között mozog egy adriai nyaralás, míg a Balaton esetében 630 ezer és 1,18 millió között alakul. Az, hogy ezeket az intervallumokat miként számolta ki Sulyok, vagy, hogy milyen módszertannal dolgozott, sajnos nem derül ki a Portfolio.hu (2024) beszámolójából.

Összességében megállapíthatjuk, hogy ugyan a szálláshelyek árának meghatározó tényezőiről számos kutatás készült, az általunk vizsgált térségben meglehetősen kevés ilyen publikált kutatást találtunk. A fentebb említett tanulmányok tudásanyagát felhasználva a következő fejezetben bemutatjuk, hogy milyen adatokat és változókat használtunk a szállásárak modellezésének érdekében.

3. Adatok

Az adatainkat úgynevezett web-scraping módszerrel töltöttük le a Booking.com¹ honlapjáról. A web-scraping módszer lényege, hogy nagy mennyiségű adatot kérdezzünk le nyilvánosan elérhető honlapokról különböző szoftveres megoldások segítségével. A Booking.com oldalon található hirdetések adatainak letöltéséhez a Python programnyelvet és az Octoparse ingyenes szoftvert használtuk segítségül. A web-scraping egyik előnye, hogy így nagy számú szállás adataihoz férünk hozzá, és minden egyes megfigyelésünk egy valójában meghirdetett szállás, amelynek a pontos jellemzőit a szállás tulajdonosa tette közzé a Booking.com oldalon. A Booking.com-ról származó adatok további előnye, hogy ez az egyik legforgalmasabb szállásfoglaló portál (Booking Holdings Inc., 2024), így magának a szállásadóknak is érdeke, hogy megjelenjenek az oldalon, különben rosszabb helyzetbe kerülnének a konkurens szállásadókkal szemben. Ez abból a szempontból fontos, hogy ezek alapján feltételezhetjük, hogy a Booking.com-on az összes foglalható szállás jelentős része megtalálható, tehát a Booking.com-ról letöltött adatok kellően nagy mintaelemszámot nyújtanak ahhoz, hogy a teljes sokaságra is érvényesek legyenek a későbbiekben levont következtetéseink.

Az általunk lekérdezett szállásadatok két felnőtt részére szóltak, a 2025. július 14. és július 20. közötti egy hétre (azaz hat éjszakára). Ez az időszak mind a két helyszínen már főszezonnak számít. A KSH (2024a) adatai szerint például júliusban 2021 és 2024 között konzisztensen ebben a hónapban volt a legnagyobb forgalma a balatoni turisztikai térségbe eső szálláshelyeknek. Az adatokat 2024. október 21-én töltöttük le. A szállásárakat amerikai dollárról átváltottuk forintba a letöltés napján az aktuális 369,88 USD/HUF árfolyam mellett².

Az 1. táblázatban láthatóak az adatbázisunkban szereplő változók. Olyan változókat válogattunk ki a letöltött adatokból, amelyek vagy a bemutatott szakirodalomban gyakran előfordulnak, vagy azokhoz hasonló jellegűek.

Változó neve	Változó leírása
Ár	6 éjszaka ára 2 felnőtt számára, forintban
Értékelés	Booking.com átlagos felhasználói értékelése, 1-10-es skálán
Távolság	vízparttól mért távolság, méterben
Medence	van-e medence az adott szálláshelyen (1 ha igen, 0 ha nem)

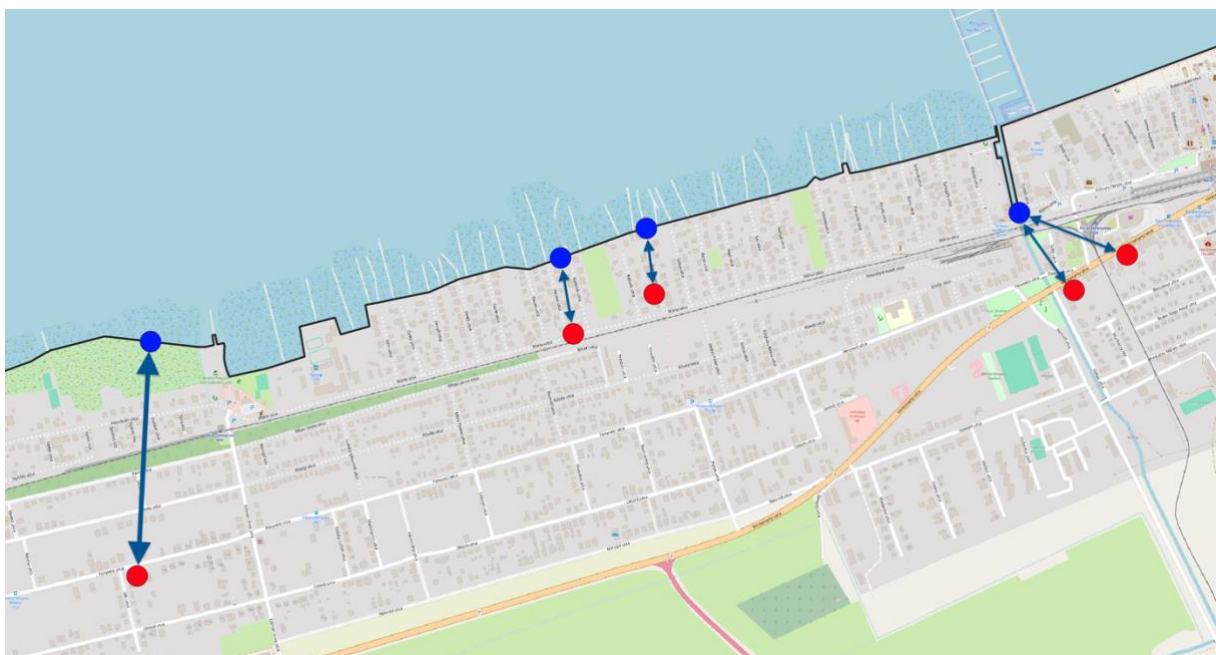
¹ Elérhető: <https://www.booking.com/>

² Elérhető: <https://www.mnb.hu/>

Takarítás	takarítanak-e naponta az adott szálláshelyen (1 ha igen, 0 ha nem)
Légkondicionálás	van-e légkondicionálás az adott szálláshelyen (1 ha igen, 0 ha nem)
Wellness	van-e wellness szolgáltatás az adott szálláshelyen (1 ha igen, 0 ha nem)
Parkolás	van-e ingyenes parkolás az adott szálláshelyen (1 ha igen, 0 ha nem)
Kilátás	van-e vízre néző kilátása az adott szálláshelynek (1 ha igen, 0 ha nem)
Hotel	szálláshely típusa (1 ha hotel, 0 ha nem)
Törölköző	van-e törölköző az adott szálláshelyen (1 ha igen, 0 ha nem)
Hajszárító	van-e hajszárító az adott szálláshelyen (1 ha igen, 0 ha nem)
Konyha	van-e konyha az adott szálláshelyen (1 ha igen, 0 ha nem)
Felszerelt	magas szobafelszereltségű a szálláshely (1 ha igen, 0 ha nem)
Ország	1 ha horvátországi a szálláshely, 0 ha balatoni

1. táblázat: Változólista, Forrás: saját szerkesztés

A *Távolság* változó azt méri, hogy az adott szálláshely milyen messzire található a legközelebbi vízparttól. Ez az adat nem szerepelt eredetileg a letöltött adatbázisban, csak a szállások koordinátái, így térinformatikai szoftver segítségével számoltuk ki azt, hogy mekkora az egyes szállások és a legközelebbi vízpart távolsága. Az *1. ábrán* látható, hogy ezzel a módszerrel kifejezetten pontos értékeket kaptunk arról, hogy az adott szállás milyen messze van a víztől. Azokat a szállásokat tartottuk meg, amelyeknek a vízparttól vett távolságuk legfeljebb 50 kilométer volt.

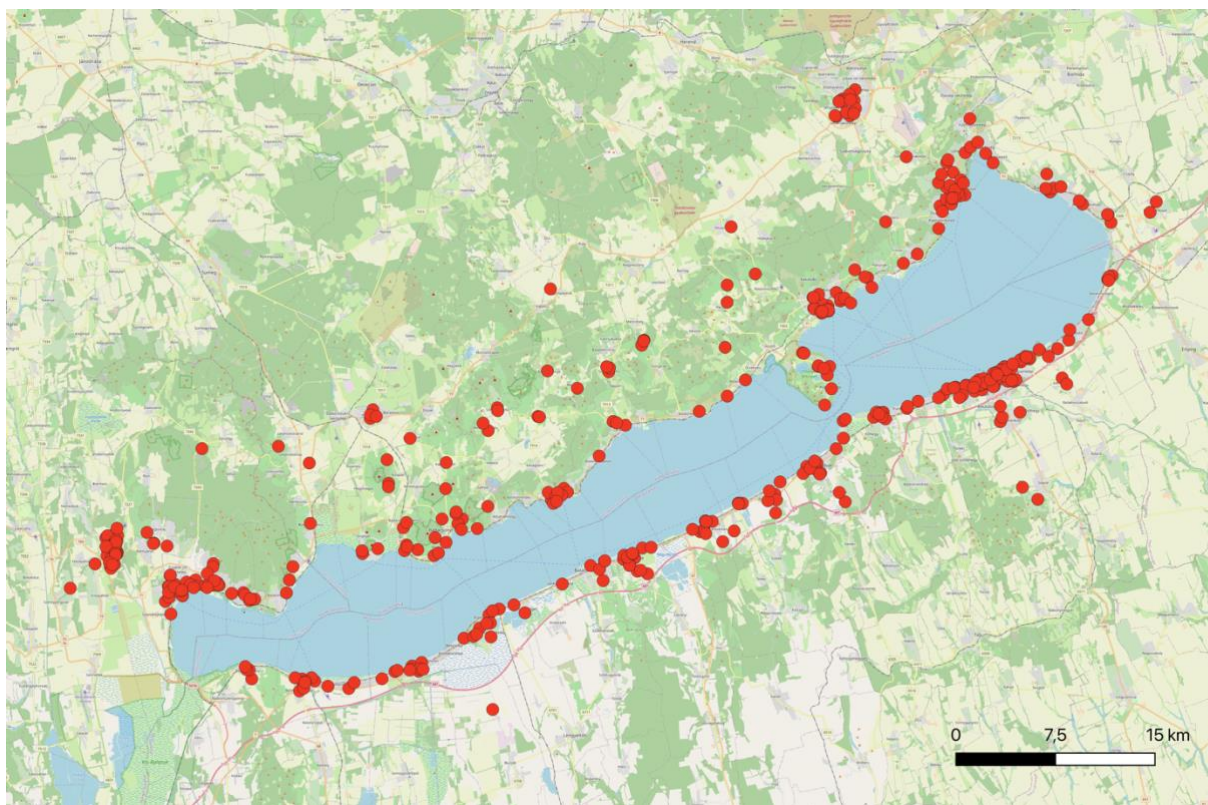


1. ábra: Távolság mérése a vízparttól, Forrás: saját szerkesztés

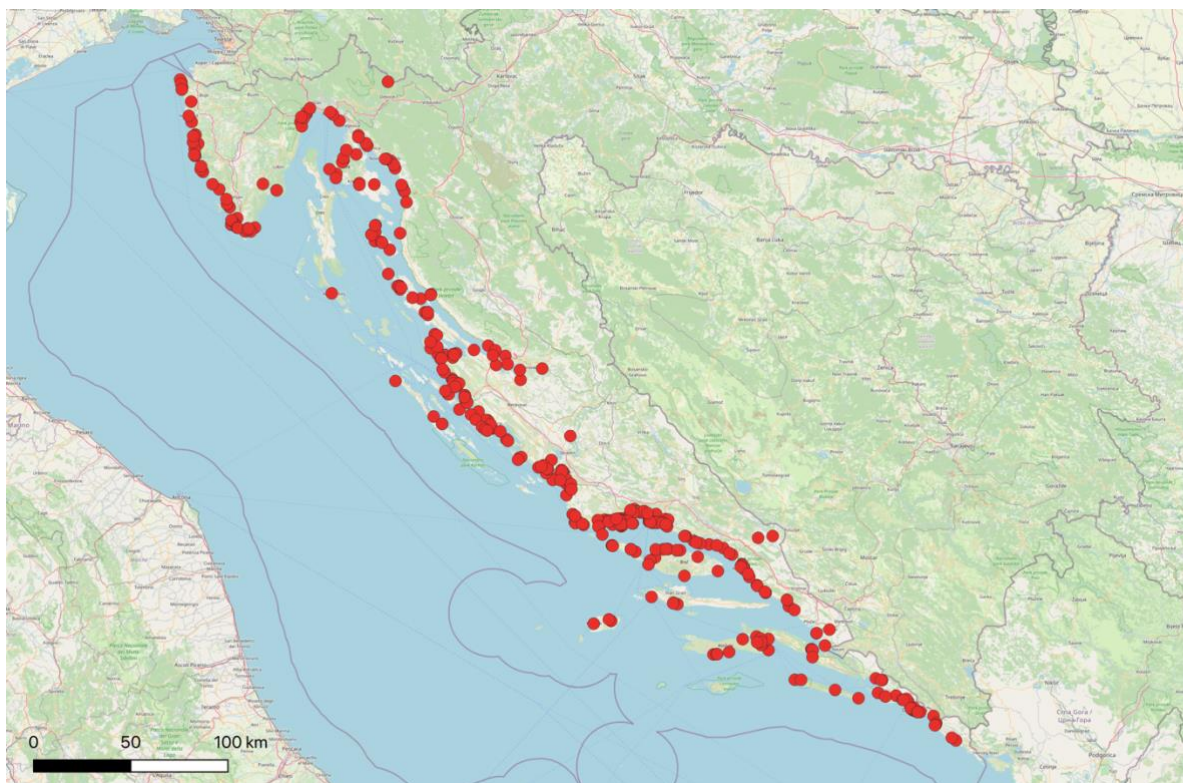
A *Parkolás* változó 1-es értéket vesz fel, hogyha van ingyenes parkolás a szálláshelyen, és 0-s értéket vesz fel, hogyha nincs. A Booking.com-on számos paramétert meg tudnak adni a szállásadók a parkolással kapcsolatban. Azokat definiáltuk ingyenes parkolásként, ahol a szállás felár nélkül kínál biztos parkolási lehetőséget (mélygarázs, belső udvar, privát parkoló).

Minden szállás esetében az adott szobára számos paraméter vonatkozott a felszereltségével kapcsolatban. Azokat a szobákat definiáltuk a *Felszerelt* változó esetében magas szobafelszereltségüként, amelyeknek a medián értéknél több kényelmi felszereltségű tulajdonsága volt.

Összesen 847 szállás adatai álltak rendelkezésünkre, ebből 406 balatoni és 441 horvát megfigyelésünk volt. A balatoni szállások földrajzi elhelyezkedése a 2. ábrán, a horvátoké a 3. ábrán látszódik.



2. ábra: Balatoni szálláshelyek az adatbázisban, Forrás: saját szerkesztés



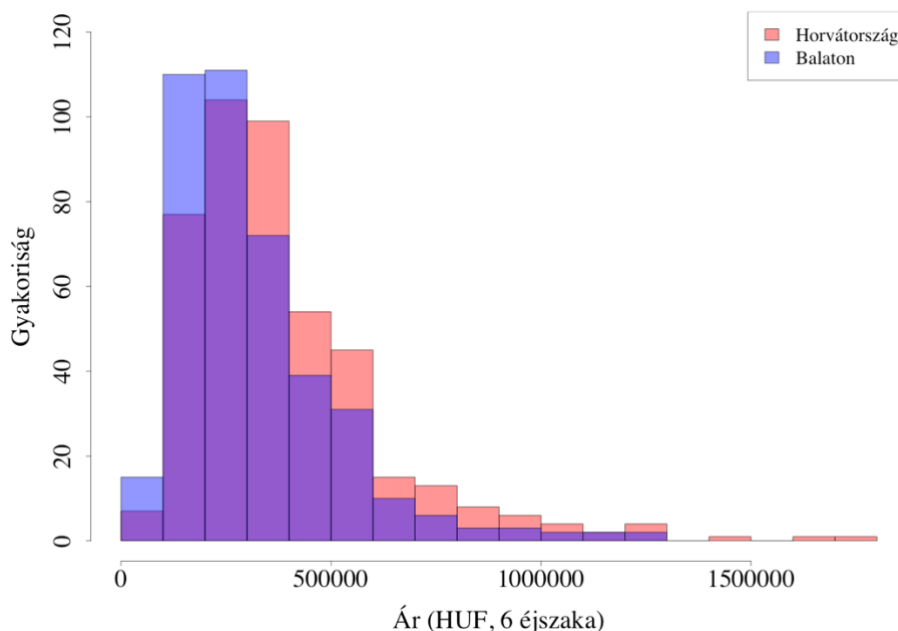
3. ábra: Horvát szálláshelyek az adatbázisban, Forrás: saját szerkesztés

A folytonos változókról szóló leíró statisztikákat a 2. táblázat tartalmazza. Látható, hogy a balatoni szállások átlagára alacsonyabb, mint a horvát szállásoké. Viszont ebből önmagában nem szabad levonni következtetéseket, mivel a szállások jellemzőjükben is eltérnek, így nem tudjuk, hogy az átlagárbeli különbség mekkora része tudható be a jellemzők eloszlásbeli különbségének a két ország között. Az is megállapítható, hogy a felhasználók dominánsan pozitív értékeléseket hagynak, relatíve alacsony szórással. Ami jelentős különbség továbbá a két helyszín között, hogy jellemzően a szállások sokkal közelebb helyezkednek el a vízparthoz Horvátországban, mint a Balatonon. Ezt a különbséget okozhatja az is, hogy a balatoni üdülőhelyek között veszprémi és hévízi szállások is vannak.

Változó	Balaton			Horvátország		
	Átlag	Medián	Szórás	Átlag	Medián	Szórás
Ár (Ft)	315 639	267 325	195 103	385 031	327 450	240 640
Értékelés	9,09	9,3	0,86	9,14	9,3	0,59
Távolság	2278	662	3 202	1211	138	6 390

2. táblázat: Folytonos változók leíró statisztikái, Forrás: saját szerkesztés

Az árak helyszín szerinti eloszlása a 4. ábrán látható. Mind a két helyen a szállásárak erősen balra ferde eloszlást követnek, viszont itt is látható, hogy a horvát szállások esetében az átlag és a medián magasabb értéket vesz fel, mint a balatoni szállásoknál.



4. ábra: A szállásárak hisztogramja, Forrás: saját szerkesztés

A kategorikus dummy változók helyszín szerinti megoszlása az 3. táblázatban látható. Ami kiemelendő, hogy a horvát szállások esetében sokkal magasabb hányadban található medence, mint a Balaton esetében. Az adriai szállások több, mint 98%-a rendelkezik légkondicionálással, míg a Balatonon ez a szám csak 82,27% volt: ez a különbség feltehetőleg betudható a melegebb éghajlatnak, amely Horvátországra jellemző. Ezeket a különbségeket leszámítva a kettő térség közel hasonló megoszlásokkal rendelkezik a kategorikus változók tekintetében.

Változó	Balaton		Horvátország	
	Igen	Nem	Igen	Nem
Medence	23,15%	76,85%	73,92%	26,08%
Hotel	10,10%	89,90%	12,47%	87,53%
Törölköző	50,99%	49,01%	65,08%	34,92%
Hajszárító	66,26%	33,74%	78,23%	21,77%
Konyha	78,33%	21,67%	80,50%	19,50%

Takarítás	2,46%	97,54%	2,95%	97,05%
Légkondi	82,27%	17,73%	98,64%	1,36%
Wellness	47,54%	52,46%	43,54%	56,46%
Parkolás	12,56%	87,44%	17,69%	82,31%
Kilátás	13,30%	86,70%	47,62%	52,38%
Felszerelt	45,81%	54,19%	49,89%	50,11%

3. táblázat: A dummy változók megoszlása országoként, Forrás: saját szerkesztés

4. Módszertan

Ebben a fejezetben bemutatjuk a kutatás során felhasznált statisztikai módszertanokat. Többváltozós lineáris regresszióval modellezzük a szállások árát, majd propensity score matching módszerrel megvizsgáljuk a hasonló jellemzőkkel rendelkező horvát és balatoni szállások közti árkülönbséget. A propensity score-okat logisztikus regressziós modellezéssel kapjuk meg. A párosított szállások árának várható értékének egyezőségéről Welch-próba alapján döntünk.

4.1. Többváltozós lineáris regresszió

Többváltozós lineáris regressziót használtunk annak érdekében, hogy a szállások árát magyarázzuk a többi megfigyelhető tulajdonsággal. A modell a változókhoz tartozó együttthatókat a legkisebb négyzetek módszerével (OLS) becsüli. Az OLS becsléssel egy minimalizációs problémát oldunk meg, mely során a modell paramétereit úgy választjuk meg, hogy a négyzetes hibák összege a lehető legkisebb legyen (Wooldridge, 2016).

Az általános egyenletet k számú magyarázó változó esetén, ahol a β együttthatók értékét keressük, az (1) egyenlet mutatja.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k \quad (1)$$

A $k + 1$ számú együttthatót úgy választjuk meg, hogy az eltérések négyzetösszege minimális legyen. A minimalizálandó kifejezés a (2) egyenlet.

$$\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \dots - \hat{\beta}_k x_{ik})^2 \quad (2)$$

Több olyan modell közül, amelyek magyarázó változóikban különböznek egymástól, az információs kritériumok segítségével tudjuk a legjobbat kiválasztani. A kettő leggyakrabban használt információs kritérium az Akaike (AIC) és a Schwarz (BIC), amelyek a modell illeszkedését a bevett magyarázó változók számának figyelembevételével ellensúlyozzák, így büntetve a modellek túlillesztését (James, 2013).

4.2. Propensity score matching

A propensity score matching módszerét leggyakrabban az orvostudományban és gazdaságpolitikai beavatkozások vizsgálatára alkalmazzák (Kane et al., 2020; Caliendo és Kopeing, 2008). Lényege, hogy kettő csoport egymáshoz leginkább hasonló egyedeit összepárosítja, majd az így kapott párok között megvizsgálható egy tetszőleges kimeneti változó értékének különbsége (Kane et al., 2020). Ez a mi esetünkben lehetővé teszi, hogy megtaláljuk az egymáshoz legjobban hasonlító balatoni és horvát szállásokat, majd megnézzük, hogy mekkora az árkülönbség az egymáshoz hasonló szállások esetén. A párosítás során létrejött szálláspárok tehát hasonló karakterisztikákkal rendelkeznek, így a párosított üdülők árának különbsége már nem fogja tartalmazni a megfigyelt változók által okozott torzítást (Rosenbaum és Rubin, 1983).

Ahhoz, hogy megtaláljuk, mely balatoni és horvát szállások hasonlítanak egymáshoz a legjobban, először ki kell számolni minden megfigyeléshez az úgynevezett propensity score-t, amely alapján össze tudjuk párosítani a későbbiekben a szállásokat. A propensity score-okat logisztikus regresszió segítségével kapjuk meg, amelyben a szállás tulajdonságaival modellezzük annak a valószínűségét, hogy az adott megfigyelés egy Horvátországban található szállás. Az így kapott becsült feltételes valószínűség lesz minden megfigyelés esetén a propensity score értéke.

4.2.1. Logisztikus regresszió

A kimeneti bináris változónk az *Ország* változó, amely akkor vesz fel 1-es értéket, ha az adott megfigyelés Horvátországban található, és 0 az értéke, hogyha balatoni. Ezt a bináris függő változót nem becsülhetjük lineáris regresszióval, mivel a függvény linearitásából következik, hogy a kiugró értékek esetén a becslés $[0,1]$ intervallumon kívüli értékeket is adna. Mivel a kimeneti változó becsült értékeit feltételes valószínűségeként szeretnénk értelmezni, ezért a 0 alatti és 1 fölötti értékeknek nincs értelme. További probléma az OLS becsléssel, hogy hibatag eloszlása nem lenne normális, vagyis léteznek kisebb hibával becsülő nemlineáris függvények (Maddala, 1983).

A folyamat első lépéseként a magyarázó változók lineáris kombinációjaként minden megfigyeléshez kapunk egy logit értéket a (3) egyenlet alapján.

$$L = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k \quad (3)$$

Ezt az értéket még nem értelmezhetjük valószínűségként, mivel nem csak a $[0,1]$ intervallumon belül vesz fel értékeket. Ahhoz, hogy az esemény bekövetkezésének valószínűségét megkapjuk, a (4) egyenletben szereplő átalakított formát kell használnunk, amely már kizárólag 0 és 1 közé becsül. Ezt az átalakítást nevezzük logit transzformációnak (Hosmer, et al., 2013).

$$P(Y = 1) = \frac{1}{1 + e^{-L}} \quad (4)$$

A β paraméterek becslése nem a legkisebb négyzetek módszerével, hanem az úgynevezett maximum likelihood módszerrel történik. A folyamat olyan β együtthatókat becsül, amelyek mellett maximális annak a valószínűsége, hogy véletlen mintavételezés esetén az adott mintát kapjuk (Ferenci, 2023). A maximum likelihood módszert az (5) egyenlet írja le.

$$L(\beta) = \prod_{y_i=1} P(y_i = 1) \prod_{y_i=0} P(y_i = 0) \rightarrow \max \quad (5)$$

4.2.2. Matching

A megfigyelések propensity score alapon történő párosítására számos különböző módszer létezik.

Ezek közül a legelterjedtebb a nearest neighbour matching, amely során az egyik csoport minden tagjához hozzárendeljük a másik csoport legközelebbi megfigyelését (Austin, 2014). A nearest neighbour matching rendelkezik egy úgynevezett caliper hiperparaméterrel, amely megad egy maximum értéket a párok propensity score-jai közötti eltérésnek, ezzel biztosítva, hogy hasonló tulajdonságú megfigyeléseket párosít össze. Az érték lehet a logit modellel becsült valószínűségek szórásához viszonyított érték vagy konkrét számszerű érték is (Wang et al., 2013). A szállások párosítására a nearest neighbour matching módszert fogjuk alkalmazni 0,1-es caliper értékkel, amely a becsült értékek szórásának 10%-át jelenti.

Egy másik módszer az optimális párosítás, amely a párok közti távolságok összegét minimalizálja, viszont ez egymástól jelentősen eltérő tulajdonságokkal rendelkező párokat is eredményezhet (Austin, 2014). Léteznek ezen kívül nemparaméteres módszerek is a

párosításra, például a kernel matching és a local linear matching, viszont ezek ritkábban alkalmazott matching módszerek számítanak (Caliendo és Kopeing, 2008).

A matchinget követően fontos a magyarázó változók értékeinek kiegyensúlyozottságát is megvizsgálni a párosított mintában. Amennyiben kiegyensúlyozottak a magyarázó változók, sikerült az általuk okozott torzításra kontrollálni a propensity score matching folyamata során. Ezt a kiegyensúlyozottságot a standardizált átlagkülönbséggel (SMD) lehet mérni, amely a két csoport közti átlagos változóértékek százalékos különbségét mutatja meg (Benedetto et al., 2018). A szakirodalomban az általánosan elfogadott küszöbérték a 10%-os SMD, amely érték alatt a változók kiegyensúlyozottnak tekinthetők (Zhang et al., 2019).

4.3. Welch-próba

A Welch-próba olyan kétmintás várhatóérték teszt, amely akkor alkalmazandó, hogyha a mintákban a szórás egyezőség feltétele nem teljesül (Welch, 1947). A teszt próbafüggvénye a (6) egyenletben olvasható, ahol $\bar{x}$ és $\bar{y}$ a mintaátlagokat, s_x és s_y a minták szórásait, n és m pedig a minták elemszámait jelentik.

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \quad (6)$$

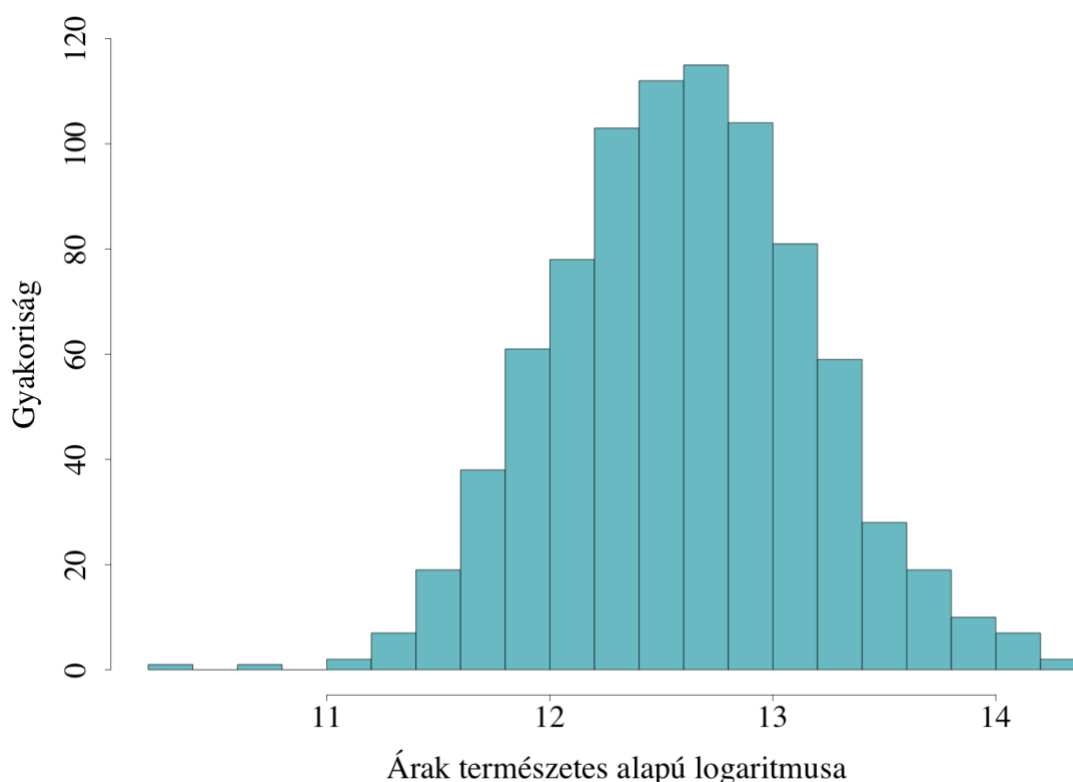
A próbafüggvény értéke t-eloszlást követ. A próba nullhipotézise azt mondja ki, hogy a két sokaság várható értéke nem tér el szignifikánsan, azonosnak tekinthető. Az alternatív hipotézis szerint a két várható érték nem tekinthető egyenlőnek.

5. Eredmények

A hedonikus árképzés modellezését lineáris regresszió segítségével végeztük annak érdekében, hogy megállapítsuk mely változók állnak szignifikáns kapcsolatban a szállások áraival. A kiválogatott változókkal logit valószínűségi modellt futtattunk, amellyel meghatároztuk az egyes szállásokhoz tartozó propensity score értékeket. Ezt követően a nearest neighbour matching módszerrel összepárosítottuk az egymáshoz leginkább hasonló szállásokat, majd Welch-próba alapján döntöttünk arról, hogy a párosított mintában szignifikánsan eltérnek-e egymástól a balatoni és horvát szállások árai.

5.1. Lineáris regressziós modell

Először négy lineáris regresszió alapuló modellt futtattunk az összes szállásra. Az első modellben a függő változó az $\hat{A}r$ volt. A második modellben a függő változó az $\hat{A}r$ logaritmus volt. Erre a transzformációra azért volt szükség, hogy az árak eloszlását szimmetrikusabbá tegyük, így pontosabb becsléseket kapjunk, mivel a lineáris regresszió a kimeneti változó átlagos értékét modellezi (Benoit, 2011). Ahogy az 5. ábrán láthatjuk, az $\hat{A}r$ logaritmusának az eloszlása közel szimmetrikus.



5. ábra: A szállásárak logaritmusának hisztogramja, Forrás: saját szerkesztés

A harmadik modellben ugyanilyen megfontolásból a *Távolságnak* is a logaritmizált értékeit vettük be. A negyedik modellben csak azok a magyarázó változók szerepelnek, amelyek a harmadik modellben szignifikánsak lettek. A modellek regressziós eredményei a 4. táblázatban tekinthetők meg.

	Függő változó:			
	Ár	log(Ár)		
	1. Modell	2. Modell	3. Modell	4. modell
Értékelés	32,031*** (10,774)	0.087*** (0.027)	0.088*** (0.027)	0.107*** (0.026)

Távolság	-5.773*** (1.421)	-0.000*** (0.000)		
ln(Távolság)			-0.085*** (0.012)	-0.093*** (0.010)
Medence	-25,328 (16,756)	-0.039 (0.042)	-0.065 (0.042)	
Hotel	91,771** (39,122)	0.348*** (0.099)	0.342*** (0.098)	0.350*** (0.069)
Törölköző	-31,271** (15,036)	-0.073* (0.038)	-0.071* (0.038)	
Hajszárító	24,261 (18,122)	0.053 (0.046)	0.060 (0.045)	
Konyha	30,456 (24,688)	0.141** (0.062)	0.190*** (0.062)	0.253*** (0.054)
Takarítás	-34,232 (46,124)	-0.023 (0.117)	-0.005 (0.116)	
Légkondicionálás	34,485 (26,227)	0.131** (0.066)	0.099 (0.066)	
Wellness	92,355*** (15,215)	0.234*** (0.038)	0.234*** (0.038)	0.250*** (0.038)
Parkolás	1,434 (20,135)	0.027 (0.051)	0.032 (0.050)	
Kilátás	46,191*** (17,685)	0.079* (0.045)	0.051 (0.045)	
Felszerelt	15,190 (17,154)	0.049 (0.043)	0.067 (0.043)	
Ország	54,891*** (17,876)	0.142*** (0.045)	0.051 (0.047)	
Konstans	-73,604 (93,485)	11.375*** (0.236)	11.865*** (0.243)	11.823*** (0.234)
N	847	847	847	847
R ²	0.161	0.199	0.211	0.195
Korrigált R ²	0.146	0.185	0.197	0.190
AIC	23146.51	1313.50	1301.27	1298.18

BIC	23227.12	1394.11	1381.88	1331.37
Reziduális standard hiba	205,735	0.520	0.516	0.518
Globális F-próba	10.62***	13.80***	14.80***	40.73***

Jelmagyarázat: 0,1%-on szignifikáns:*** 1%-on szignifikáns: ** 5%-on szignifikáns: *

4. Táblázat: Lineáris regressziós modellek, Forrás: saját szerkesztés

Az első modellt a lineáris eredményváltozó ferdesége, így a nem megfelelő függvényforma miatt elvetettük. A többi modell közül az Akaike és Schwarz információs kritérium alapján választottuk ki a legjobbat, mind a kettő egyértelműen a negyedik modellt preferálta. A konstans tagot leszámítva 5 szignifikáns változót kaptunk. Az összes szignifikáns változó együtthatójának előjele megfelel az előzetes várakozásainknak. Azt kaptuk, hogy minél jobb egy hotel értékelése a Booking.com-on, várhatóan annál magasabb a szállás ára. Amennyiben az adott szállás egy hotel, vagy rendelkezik konyhával vagy wellness szolgáltatással, várhatóan többet kell fizetni az üdülőhelyért. Az is látszik a regresszió eredményeiből, hogy minél messzebb van egy szállás a vízparttól, ceteris paribus annál kevesebbet kell fizetni az eltöltött éjszakákért. Ami érdekes, hogy maga az *Ország* változó nem lett szignifikáns, vagyis úgy tűnik előzetesen, hogy az, hogy egy adott szállás a Balatonon vagy Horvátországban található, a többi változó változatlansága mellett nem jelent szignifikáns különbséget az ár tekintetében.

A végső modellünk így a negyedik modell lett, amelynek felírása a (7) egyenletben olvasható. Ezt a modellt külön-külön lefuttattuk a két helyszínre, és mind a két esetben az összes magyarázó változó szignifikáns volt és az együtthatók előjele is megegyezett. Ebből arra következtethetünk, hogy mindkét országban a vizsgált változók szignifikáns kapcsolatban állnak a szállások áraival.

$$\ln(\widehat{Ár}) = \beta_0 + \beta_1 \text{Értékelés} + \beta_2 \ln(\text{Távolság}) + \beta_3 \text{Hotel} + \beta_4 \text{Konyha} + \beta_5 \text{Wellness}$$

(7)

5.2. Propensity score kiszámítása logit modellel

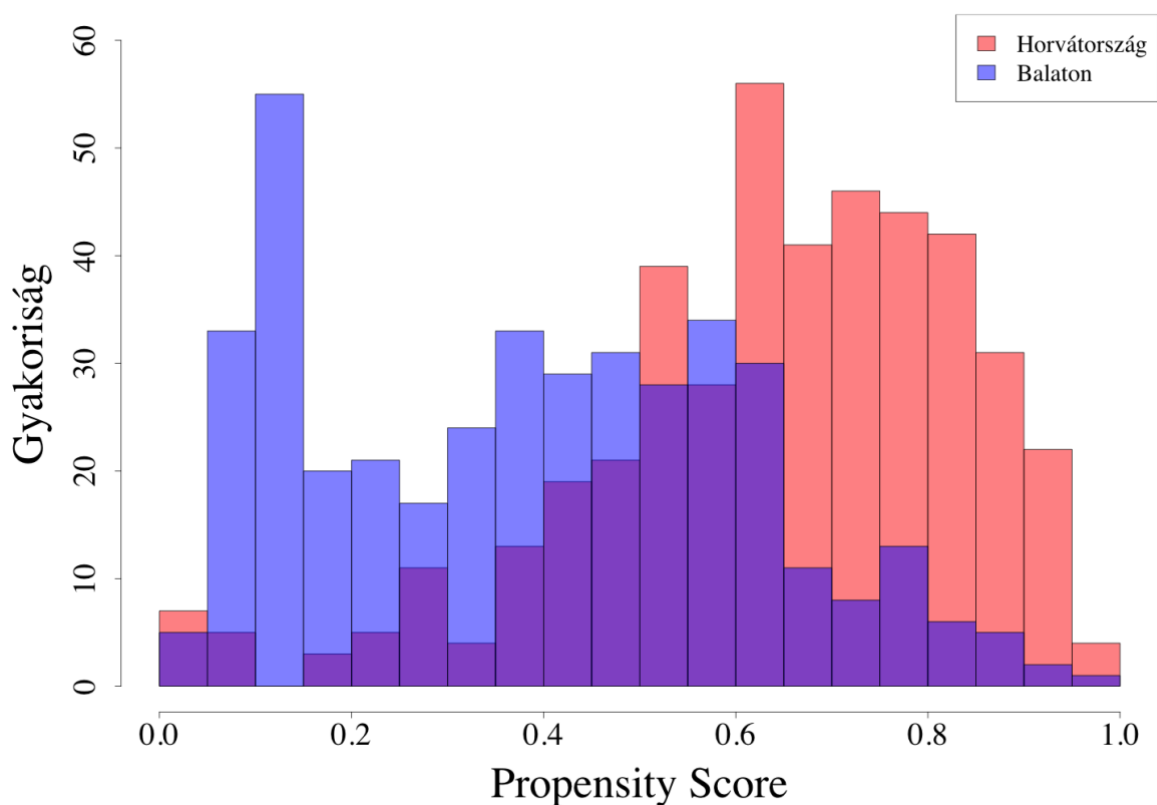
Az egyes szállásokhoz tartozó propensity score-ok kiszámításához logit modellt használtunk. A függő változó az *Ország*, azaz ezzel a modellel gyakorlatilag azt becsüljük meg, hogy mennyi a valószínűsége annak, hogy az adott szállás Horvátországban található. Azokat a magyarázó változókat tettük bele ebbe a modellbe, amelyek a (7) egyenletben szerepelnek. A mögöttes intuíció a változók választása mögött az volt, hogy azokat vettük be a logit modellbe, amelyekről már a lineáris regresszió segítségével beláttuk, hogy szignifikáns kapcsolatban állhatnak a szálláshelyek árával. A logit modell eredményei az 5. táblázatban olvashatóak, ahonnan látható, hogy a logit modellben is minden magyarázó változó szignifikáns lett.

	<i>Függő változó:</i>
	<i>Ország</i>
Értékelés	0.194 (0.115)*
ln(Távolság)	-0.737 (0.061)***
Hotel	0.536 (0.316)*
Konyha	0.560 (0.248)**
Wellness	-0.290 (0.167)*
Konstans	2.215 (1.041)**
N	847
Log Likelihood	-473.998

Jelmagyarázat: 0,1%-on szignifikáns:*** 1%-on szignifikáns: ** 5%-on szignifikáns: *

5. táblázat: Logit modell, Forrás: saját szerkesztés

A logit modell által becsült propensity score-ok eloszlását ország szerinti bontásban a 6. ábrán láthatjuk. A modell becsült valószínűségek eloszlása a két ország esetén erősen eltér, amely abból fakad, hogy a Balaton és Horvátország között a bevett magyarázó változók eloszlása jelentősen különbözik. Ez tehát azt jelenti, hogy a szállások jellemzői és karakterisztikái eltérőek a két hely között, így fontos ezekre az eloszlásbeli különbségekre kontrollálunk, hogy az árbeli különbségeket az összetétel hatást kiszűrve tudjuk megvizsgálni.



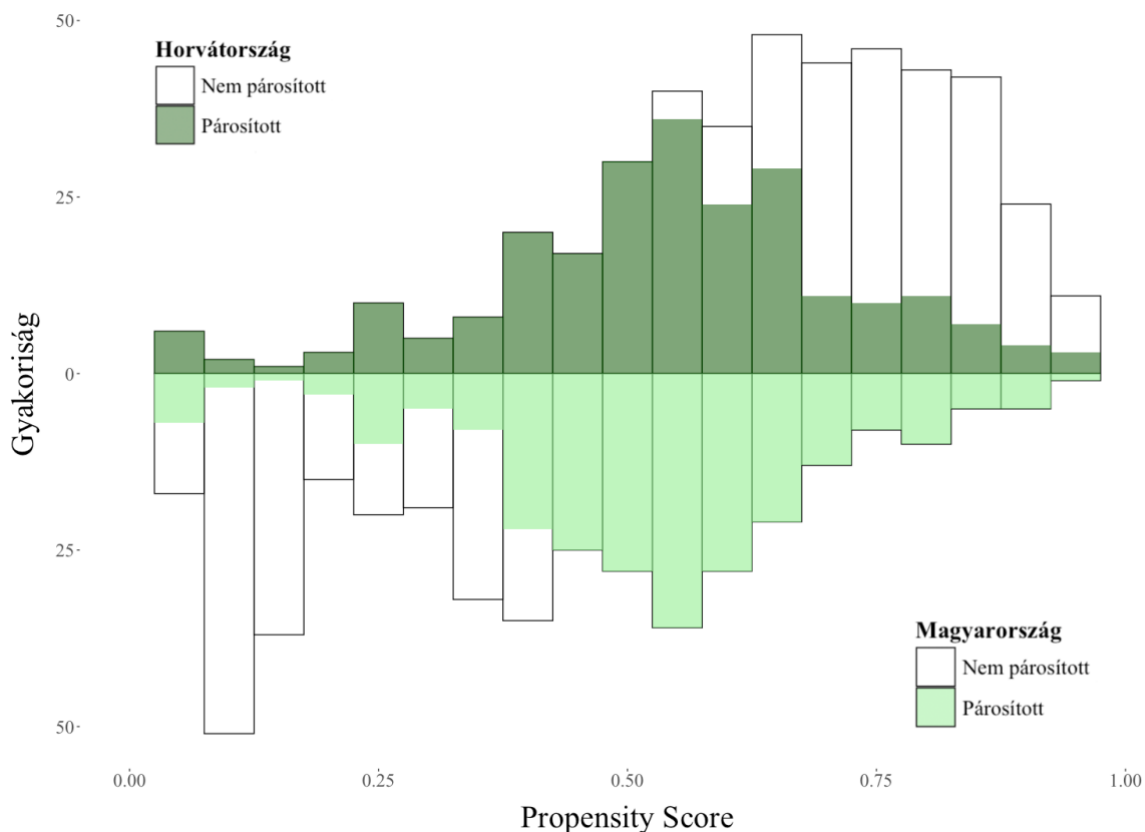
6. ábra: A propensity score-ok eloszlása országonként, Forrás: saját szerkesztés

5.3. Matching

A propensity score-ok kiszámítása után el tudjuk végezni a párosítást, amelynek az a célja, hogy a 6. ábrán látható eloszlásbeli különbségeket kiegyenlítse. A matchinget a nearest neighbour matching módszerével végeztük el, 0,1-es caliper hiperparaméter beállítással. Így a módszer azokat a szállásokat párosította össze, amelyeknek a propensity score-ja egymástól kevesebb, mint a propensity score-ok szórásának 10%-ával tért el. A szakirodalomban leggyakrabban a 0,2-es caliper értéket alkalmazzák (Wang et al., 2013; Baek et al., 2015), viszont kutatásunkban azért a 0,1-es küszöbértéket választottuk, hogy az egymáshoz legjobban hasonló egyedek kerüljenek egy párba (Lunt, 2014). A caliper értékének csökkentése azzal jár, hogy egymáshoz jobban hasonló megfigyeléseket párosítunk össze, viszont emiatt összesen több szállás marad ki a matchingből. Azért a 0,1-es caliper értéket választottuk, mivel célunk a pontosabb párosítás volt, és a 0,1-es érték a 0,2-eshez képest csak 15-tel kevesebb párt hozott létre.

A matching eljárást követően 238 szálláspárt kaptunk, vagyis a matching nem járt jelentős adatvesztéssel, mivel a megfigyelések jelentős részét sikerült párokba rendezni. A

propensity score-ok párosítás előtti és utáni eloszlása a 7. ábrán található meg. A propensity score-ok eloszlásai kiegyenlítődtek a matching eljárást követően az ábra alapján, tehát első ránézésre előzetes célunk teljesült, kontrolláltunk az eloszlásbeli különbségekre.

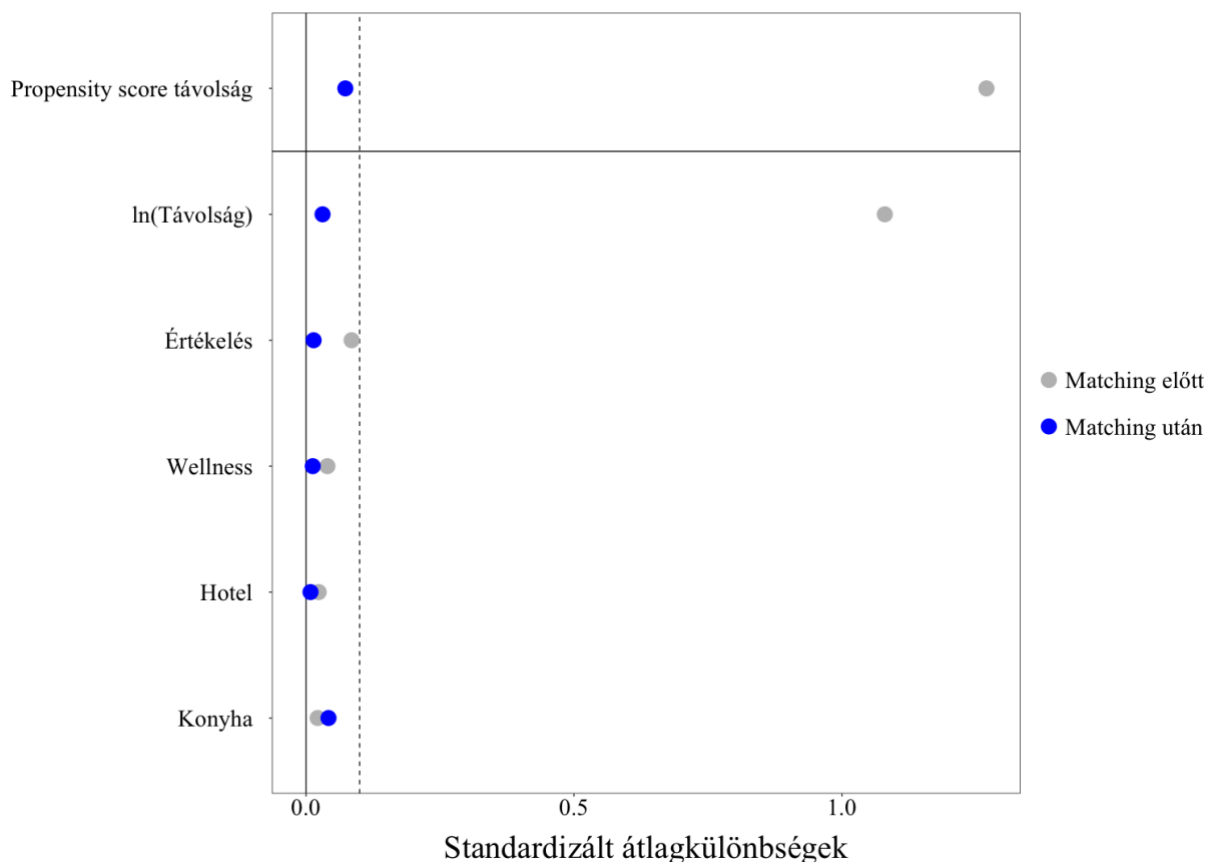


7. ábra: A propensity score-ok eloszlása a teljes és a párosított mintában, országonként

Forrás: saját szerkesztés

Ahhoz, hogy meggyőződjünk arról, hogy az eljárás során valóban hasonló tulajdonságokkal rendelkező szállásokat sikerült összepárosítanunk, a magyarázó változók értékeinek eloszlásait külön-külön is ellenőriznünk kell. Ehhez a változók kiegyensúlyozottságát kell megvizsgálnunk, amelyek akkor tekinthetők kiegyensúlyozottnak, hogyha az SMD értéke nem haladja meg a 10%-ot (Zhang et al., 2019). Az SMD változónkénti értékeit a 8. ábrán láthatjuk. Ez alapján egyértelmű, hogy míg a matching eljárás előtt a *Távolság* logaritmus a kiegyensúlyozatlan volt, a párosítást követően a változó standardizált átlagkülönbsége 10% alatti lett, tehát a párosított szállások átlagos vízparttól mért távolságának logaritmus a két csoport között kiegyensúlyozottá vált. A többi bevett magyarázó változó esetében azt találjuk, hogy eloszlásuk a párosítás előtt és után is kiegyensúlyozott volt. Ebből arra következtethetünk, hogy a szállások karakterisztikái leginkább abban tértek el a Balaton és

Horvátország között, hogy a balatoni üdülők parttól mért távolsága jellemzően sokkal nagyobb volt, mint a horvátok esetében.



8. ábra: A standardizált átlagkülönbségek, Forrás: saját szerkesztés

5.4. Kétmintás várható érték próba

Miután a matchinggel kontrolláltunk a szállások jellemzőinek eltérő jellegére, a párosított szállások árainak különbségét egyszerű kétmintás várható érték próbával tudjuk tesztelni. Ezt azért tehetjük meg, mivel a matching segítségével kiszűrtük a jellemzők eloszlásának különbözőségéből fakadó eltéréseket (összetétel hatás), tehát a párosított balatoni és horvát szállások átlagos árkülönbsége a hasonló tulajdonságú szállások tényleges árkülönbségét mutatja. Ahhoz, hogy leteszteljük ezen különbség statisztikailag valóban szignifikáns-e, kétmintás várható érték próbát alkalmazhatunk.

A párosított mintában a balatoni és adriai szállások szórás egyezőségéről F-próba alapján döntöttünk, amelynek nullhipotézisét minden szokásos szignifikancia szinten elutasítottuk, tehát nem egyeztek meg a szórások. Mivel a szórások nem voltak azonosnak tekinthetők, így a várható értékek egyenlőségének tesztelésére Welch-próbát alkalmaztunk. A párosított mintában

2 személy részére 6 éjszaka átlagosan a Balatonnál 348 613 Ft-ba, míg Horvátország esetén 385 030 Ft-ba került. A Welch-próba próba p-értéke 7,11% lett, vagyis nem tudtuk elutasítani a nullhipotézist, miszerint a kettő hely szállásárai megegyeznek a sokaságban.

Ez alapján arra az eredményre jutottunk, hogy a hasonló jellemzőkkel rendelkező balatoni és adriai szállások árai nem térnek el egymástól szignifikánsan. Más szóval ez azt jelenti, hogy a szállás jellemzőit ugyanolyan mértékben árazzák a Balatonon és Horvátországban, és a két helyszín közötti árkülönbségek abból fakadnak elsősorban, hogy Horvátországban több olyan adottsággal rendelkező szállás található, amelyek növelik az adott szállás árát. Ezen eredmények fényében az előzetes hipotézisünket el tudjuk utasítani, miszerint a horvát Adriai-tenger menti szálláshelyek drágábbak, mint a balatoni szállásarak. Ugyanis azt találtuk, hogy a párosítást követően az egymáshoz hasonló jellemzőkkel rendelkező balatoni és horvát szállások árában nincs szignifikáns különbség.

6. Kutatás limitációi

A kutatásunk limitációi elsősorban az adatszerzés módjából és az alkalmazott módszertan jellegéből adódnak.

A vizsgált szállásárakat a Booking.com-ról töltöttük le, tehát csak azokat a szállásokat vizsgáltuk, amelyek a Booking.com-on fellelhetőek. Mivel a letöltött árak eredetileg amerikai dollárban voltak megadva, amelyeket átváltottunk forintba, így az eredményeket az árfolyamok hatása befolyásolhatja. A letöltött szállásadatok a honlap jellegéből fakadóan vagy hotelek vagy apartmanok, tehát a P2P elven alapuló szállásokat, mint például az Airbnb, nem tartalmazzák. Az adatszerzést a szállás igénybevétele előtt több, mint 8 hónappal végeztük, így dolgozatunk eredményei a last-minute foglalásokra nem vonatkoznak. Elképzelhető, hogy szignifikáns árkülönbséget találnánk a Balaton és Horvátország között, hogyha 1 héttel a szállás tervezett elfoglalása előtti árakat vizsgálnánk, viszont ez a kérdés túlmutat jelen kutatásunk keretein. A szállások árait továbbá kizárólag 2 felnőtt esetén néztük, a családos, egyszemélyes, vagy nagycsoportos szállásárakat figyelmen kívül hagytuk.

A propensity score matching hátulütője, hogy az össze nem párosított adatok miatt megfigyeléseket veszítünk az eredeti mintához képest. Továbbá a módszertan másik hátránya, hogy csak a megfigyelt változók közötti különbségeket képes csökkenteni a csoportok között. Így nem tudtunk kontrollálni azokra a változókra, amelyek nem képezték részét az adatbázisunknak, de feltehetően hatással lehetnek egy adott szállás árára. Ilyen kihagyott tényező lehet például a szállás szobamérete, a tisztaság mértéke, az étkeztetés megléte, az

üdülőhely és a település központjától vett távolsága, a szállás környezetének nyugalma, vagy a környéken található éttermek és látnivalók minősége és közelsége.

További limitáció, hogy dolgozatunkban egyedül a szállások árát hasonlítottuk össze. Így egy nyaralás egyéb anyagi vonzatait, mint például az úti költséget, vagy az ételek és szolgáltatások árát figyelmen kívül hagytuk.

7. Összefoglalás

Kutatásunk célja az volt, hogy összehasonlítsuk a balatoni és a horvát tengerparti szálláshelyek árait. A Balaton és az Adria mind turisztikai, mind gazdasági szempontból kiemelt célpontok, így a szállások folyamatos emelkedésének köszönhetően Horvátország egyre gyakrabban merül fel alternatívaként a Balatonnal szemben. Az eddigi kutatások viszont jellemzően a szállások árainak összehasonlításával nem foglalkoztak, vagy statisztikai módszereket nem alkalmaztak. Dolgozatunkban arra a kérdésre szerettünk volna átfogó választ adni a propensity score matching módszerével, hogy vajon drágábbak-e a horvát szállások, mint a balatoniak.

Elemzésünkben a szálláshelyek árait és egyéb jellemzőit a Booking.com-ról gyűjtöttük, az adatokat automatizált web-scraping módszer segítségével töltöttük le. Összesen 847 szálláshelyet vizsgáltunk kutatásunk során, amelyből 406 balatoni, 441 pedig horvát tengerparti szállás volt, és a foglalások 2025 júliusára vonatkoztak. Olyan változókat használtunk, amelyek bizonyítottan hatással vannak a szállások áraira, például a szállás felhasználói értékelése, a szoba felszereltsége, a wellness szolgáltatás megléte, a vízparttól vett távolság, az ingyenes parkolás lehetősége, vagy a légkondicionálás jelenléte.

Ahhoz, hogy meghatározzuk mely jellemzők állnak szignifikáns kapcsolatban a szállás árával, annak logaritmusára többváltozós lineáris regressziós modelleket építettünk. A végső modell alapján megállapítottuk, hogy a vízparttól vett távolság, a felhasználói értékelés, a wellness szolgáltatás, a szállás típusa (hotel vagy sem), valamint a konyha megléte szignifikáns kapcsolatban állnak a szállások árával. Ezt követően ezekkel a változókkal, mint magyarázó változókkal logit modellt futtatunk, amellyel minden szálláshoz hozzárendeltünk egy propensity score értéket. A propensity score-ok alapján a nearest neighbor matching módszerrel összepárosítottuk az egymáshoz leginkább hasonló tulajdonságokkal rendelkező balatoni és horvát szállásokat. Az így kapott 238 pár esetében kontrolláltunk a bevett változók általi eloszlásbeli különbségekre (összetétel hatás), így a változók szempontjából kiegyensúlyozott

párokat kaptunk. A párosított szállásokat Welch-próbával vizsgáltuk meg, hogy szignifikánsan eltérnek-e a balatoni és adriai szállások árának várható értékei.

Mivel a Welch-próba szerint minden szokásos szignifikanciaszinten megegyeztek a várható értékek, így arra a következtetésre jutottunk, hogy a hasonló jellemzőkkel rendelkező balatoni és horvát szállások árai között nincs statisztikailag szignifikáns különbség. Ez azt jelenti, hogy a szálláshelyek közti alapvető árkülönbségeket nem az eltérő ország, hanem a különböző szállásjellemzők eredményezik, mint például a konyha megléte, wellness szolgáltatás, vagy a vízparthoz közelebbi elhelyezkedés. Az előzetes hipotézisünket elutasítottuk, miszerint a horvát Adriai-tenger menti szálláshelyek drágábbak, mint a balatoni szállásárak. Az országok közötti árkülönbségek mögött leginkább az áll, hogy Horvátországban több a magas felszereltséggel, illetve vízparthoz közelebbi elhelyezkedéssel rendelkező szálláshely található.

Dolgozatunkat viszont a jövőben számos módon lehetne továbbfejleszteni. Érdekes lehetne még tágabb adatokat alapul venni, amelyek egyfelől több jellegű szállást foglalnának magukba (például Airbnb), másfelől több, ár szempontjából releváns változót tartalmaznának. Továbbá kutatásunkban a balatoni és horvát szállások közötti árkülönbségeket keresztmetszeti szempontból elemeztük, viszont érdekes lenne megvizsgálni az árkülönbség mértékét visszamenőleg idősoros adatokon is.

8. Irodalomjegyzék

24.hu (2024): *Balaton vagy Adria: súlyos árkülönbséget találtak a turizmuskutatók*
<https://24.hu/fn/gazdasag/2024/05/29/balaton-adria-turizmus-nyaralas-idegenforgalom-siofok-balatonfured-udules-szallas/>

Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in medicine*, 33(6), 1057-1069.

Baek, S., Park, S. H., Won, E., Park, Y. R., & Kim, H. J. (2015). Propensity score matching: a conceptual review for radiology researchers. *Korean journal of radiology*, 16(2), 286-296.

Benedetto, U., Head, S. J., Angelini, G. D., & Blackstone, E. H. (2018). Statistical primer: propensity score matching and its alternatives. *European Journal of Cardio-Thoracic Surgery*, 53(6), 1112-1117.

Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics, London*, 22(1), 23-36.

Booking Holdings Inc. (2024). *2023 Annual Report*. U.S. Securities and Exchange Commission.
<https://www.sec.gov/ix?doc=/Archives/edgar/data/1075531/000107553124000014/bkng-20231231.htm>

Booking.com. (n.d.). *Booking.com | Official site | The best hotels, flights, car rentals & accommodations*. [Adatok forrása] Letöltve: 2024, október 21. Elérhető:
<https://www.booking.com/>

Bull, A. O. (1994). Pricing a motel's location. *International Journal of Contemporary Hospitality Management*, 6(6), 10-15.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1), 31-72.

Dedeoğlu, B. B., Van Niekerk, M., Küçükergin, K. G., De Martino, M., & Okumuş, F. (2020). Effect of social media sharing on destination brand awareness and destination quality. *Journal of Vacation Marketing*, 26(1), 33-56.

Domonkos, Á., Sinkovics, K., & Retz, T. (2016). Turizmusgazdaság a Balaton idegenforgalmi régióban. *Területi Statisztika*, 56(3), 346-368.

Dudás, Gy., Boros, L., & Vida, Gy. (2020). Az Airbnb szállásait meghatározó tényezők a Balaton turisztikai régióban.

Ferenci Tamás (2023): *Regressziós modellezés*. Letöltve: 2024.11.04. Elérhető: <https://www.medstat.hu/oktatas/RegressziosModellezes/FerenciTamas-RegressziosModellezes.pdf>

Gerdesics, V. (2017). Horvátország (ország) imázsa az EU-csatlakozás időszakában: Balkán vagy Európa?

Hartman, R. S. (1989). Hedonic methods for evaluating product design and pricing strategies. *Journal of Economics and Business*, 41(3), 197-212.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.

Hung, W. T., Shang, J. K., & Wang, F. C. (2010). Pricing determinants in the hotel industry: Quantile regression analysis. *International Journal of Hospitality Management*, 29(3), 378-384.

Israeli, A. A. (2002). Star rating and corporate affiliation: their influence on room price and performance of hotels in Israel. *International Journal of Hospitality Management*, 21(4), 405-424.

James, G. (2013). An introduction to statistical learning.

Kane, L. T., Fang, T., Galetta, M. S., Goyal, D. K., Nicholson, K. J., Kepler, C. K., ... & Schroeder, G. D. (2020). Propensity score matching: a statistical method. *Clinical spine surgery*, 33(3), 120-122.

Központi Statisztikai Hivatal (2024a): *A turisztikai szálláshelyeken eltöltött külföldi és belföldi vendégéjszakák Budapesten és turisztikai térségeként, havonta* (27.2.1.17.) [Adattábla] https://www.ksh.hu/stadat_files/tur/hu/tur0081.html

Központi Statisztikai Hivatal (2024b): *A külföldre tett utazások főbb mutatói fontosabb célszágok szerint, negyedévente* (27.2.1.8.) [Adattábla] https://www.ksh.hu/stadat_files/tur/hu/tur0070.html

Lunt, M. (2014). Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching. *American journal of epidemiology*, 179(2), 226-235.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics* (Vol. 149). Cambridge University Press.

Magyar Nemzeti Bank. (n.d.). Árfolyamok. <https://www.mnb.hu/arfolyam-lekerdezes>

Moreno-González, A. A., León, C. J., & Fernández-Hernández, C. (2020). Health destination image: The influence of public health management and well-being conditions. *Journal of Destination Marketing & Management*, 16, 100430.

Pénzcentrum (2024a): *Brutál áremeléssel robban a szezon a Balatonnál: ennyivel lehet drágább a nyaralás 2024-ben* <https://www.penzcentrum.hu/utazas/20240423/brutal-aremelessel-robban-a-szezon-a-balatonnal-28-szazalekkal-is-dragabb-lehet-a-nyaralas-2024-ben-1149912>

Pénzcentrum (2024b): *Nyaralás az Adrián vagy a Balatonon: nem hiszed el, melyik a drágább 2024-ben* <https://www.penzcentrum.hu/utazas/20240531/nyaralas-az-adrian-vagy-a-balatonon-nem-hiszed-el-melyik-a-dragabb-2024-ben-1151624#>

Portfolio.hu (2024): *Balaton vs. Adria: kiszámolták, hol éri meg jobban nyaralni*
<https://www.portfolio.hu/gazdasag/20240502/balaton-vs-adria-kiszamoltak-hol-eri-meg-jobban-nyaralni-684005>

Rácz, Sz. (2016). Horvátország térszerkezete. *Tér és Társadalom*, 30(3), 81-104.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1), 34-55.

Rosenbaum, Paul R. & Rubin, Donald B. (1983): The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55.

RTL.hu (2024): *Kiderült, hol olcsóbb nyaralni, Horvátországban vagy a Balatonon*
<https://rtl.hu/gazdasag/2024/05/29/nyaralas-adria-horvatorszag-balaton-olcsobb>

Sulyok, J., & Papp, Z. M. (2023). A Balaton mint turisztikai desztináció imázsa= Image of Lake Balaton as a Tourism Destination. *COMITATUS: ÖNKORMÁNYZATI SZEMLE*, 33(245), 37-52.

Telex.hu (2024. augusztus 1.): *Leteszteltük, hogy tényleg olcsóbb-e az Adria, mint a Balaton* [Videó] <https://www.youtube.com/watch?v=tVvc7y2IcuA>

Thrane, C. (2007). Examining the determinants of room rates for hotels in capital cities: The Oslo experience. *Journal of revenue and Pricing Management*, 5, 315-323.

Wang, Y., Cai, H., Li, C., Jiang, Z., Wang, L., Song, J., & Xia, J. (2013). Optimal caliper width for propensity score matching of three treatment groups: a Monte Carlo study. *PloS one*, 8(12), e81045.

Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, 34(1-2), 28-35.

Wooldridge, J. M. (2016): Introductory econometrics: A modern approach. *Cengage Learning*. Boston.

Wu, C. W. (2016). Destination loyalty modeling of the global tourism. *Journal of business research*, 69(6), 2213-2219.

Zhang, Z., Kim, H. J., Lonjon, G., & Zhu, Y. (2019). Balance diagnostics after propensity score matching. *Annals of translational medicine*, 7(1).