

Keresztábra elemzés feladatok

1. feladat

A repter.sav állomány egy repülőtér közelében eladásra kínált ingatlanok véletlen, reprezentatív mintáját tartalmazza. A környék zajterhelésének és az ingatlan állapotának kapcsolatát vizsgálva válaszoljon SPSS segítségével az alábbi kérdésekre.

A) Írassa ki zajszintet és az állapotot tartalmazó kombinációs táblát!

állapot * Zajszint (dB) Crosstabulation

Count		Zajszint (dB)			Total
		55 dB alatt	55-59 dB	60 dB-	
állapot	felújítást igényel	93	74	23	190
	felújított	245	149	40	434
	új építésű	119	66	77	262
	újszerű	126	76	23	225
Total		583	365	163	1111

B) Írassa ki az állapot szerinti megoszlást a különböző zajzónákban külön-külön!

állapot * Zajszint (dB) Crosstabulation

		Zajszint (dB)			Total	
		55 dB alatt	55-59 dB	60 dB-		
állapot	felújítást igényel	Count	93	74	23	190
		% within Zajszint (dB)	16,0%	20,3%	14,1%	17,1%
	felújított	Count	245	149	40	434
		% within Zajszint (dB)	42,0%	40,8%	24,5%	39,1%
	új építésű	Count	119	66	77	262
		% within Zajszint (dB)	20,4%	18,1%	47,2%	23,6%
	újszerű	Count	126	76	23	225
		% within Zajszint (dB)	21,6%	20,8%	14,1%	20,3%
Total		Count	583	365	163	1111
		% within Zajszint (dB)	100,0%	100,0%	100,0%	100,0%

A táblázat alapján a mintázatok kereséséhez érdemes a zajzónákon belüli arányokat a teljes mintában megjelenő arányokhoz (utolsó oszlop) viszonyítani. Kiemelhető, hogy

- az 55 dB alatti zajzónán belül a felújított ingatlanok aránya (42,0%) valamelyest magasabb a teljes mintában megjelenő arányukhoz (39,1%) viszonyítva,
- az 55-59 dB-es zajzónán belül a felújítást igénylő ingatlanok aránya (20,3%) valamelyest magasabb a teljes mintában megjelenő arányukhoz (17,1%) viszonyítva,
- a 60 db feletti zajzónában lévő ingatlanok között tipikusnak tekinthetők az új építésű ingatlanok: az új építésű ingatlanok aránya (47,2%) kétszer akkora a 60 dB feletti zajzónán belül, mint a teljes mintán belül (23,6%).

- C) Írassa ki a függetlenség esetén fennálló gyakoriságokat és számítsa ki a kapcsolatmérő mutatót (Cramer-együttható)!

állapot * Zajszint (dB) Crosstabulation

		Zajszint (dB)			Total	
		55 dB alatt	55-59 dB	60 dB-		
állapot	felújítást igényel	Count	93	74	23	190
		Expected Count	99,7	62,4	27,9	190,0
		% within Zajszint (dB)	16,0%	20,3%	14,1%	17,1%
	felújított	Count	245	149	40	434
		Expected Count	227,7	142,6	63,7	434,0
		% within Zajszint (dB)	42,0%	40,8%	24,5%	39,1%
	új építésű	Count	119	66	77	262
		Expected Count	137,5	86,1	38,4	262,0
		% within Zajszint (dB)	20,4%	18,1%	47,2%	23,6%
	újszerű	Count	126	76	23	225
		Expected Count	118,1	73,9	33,0	225,0
		% within Zajszint (dB)	21,6%	20,8%	14,1%	20,3%
Total	Count	583	365	163	1111	
	Expected Count	583,0	365,0	163,0	1111,0	
	% within Zajszint (dB)	100,0%	100,0%	100,0%	100,0%	

Látható a táblázat alapján, hogy a függetlenség esetén fennálló gyakoriságok eltérnek a gyakoriságoktól. Kérdés azonban, hogy mekkora ez az eltérés, milyen erősségű kapcsolatról beszélhetünk?

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,239	,000
	Cramer's V	,169	,000
N of Valid Cases		1111	

A kapcsolat erősségét tekintve megállapítható, hogy a zajszint és az ingatlanok állapota között gyenge kapcsolat van (C=0,169).

- D) Ötszázalékos szignifikanciaszint mellett szignifikáns kapcsolat van-e a zajszint és az állapot között?

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	63,327 ^a	6	,000
Likelihood Ratio	56,658	6	,000
Linear-by-Linear Association	,639	1	,424
N of Valid Cases	1111		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 27,88.

A vizsgálat nullhipotézise szerint az állapot és a zajszint között nincs szignifikáns kapcsolat. A táblázat alatti megjegyzés alapján a próbafüggvény alkalmazási feltétele a cellák több mint 20%-nál teljesült, így a teszt alkalmazási feltétele teljesült. Ötszázalékos szignifikanciaszinten a nullhipotézist (sig<0,05, $\chi^2=63,327$) elvetjük, azaz a zajszint és az állapot között kimutatható szignifikáns, gyenge (C=0,169) kapcsolat

Amennyiben célunk a minta alapján az alapsokaságra való következtetés, akkor az alábbi **összefüggő szöveges elemzést** építhetjük fel (az alkalmazási feltételt kielégítő megoldásra):

Mivel egy nominális és egy ordinális mérési szintű változó közötti kapcsolatot vizsgálunk, ezért keresztábra elemzést használhatunk a probléma vizsgálatára. A vizsgálat nullhipotézise szerint az állapot és a zajszint között nincs szignifikáns kapcsolat. A zajszint és az állapot között ötszázalékos szignifikanciaszinten kimutatható szignifikáns ($\text{sig} < 0,05$, $\chi^2 = 63,327$), gyenge ($C = 0,169$) kapcsolat. Kiemelhető, hogy

- az 55 dB alatti zajzónán belül a felújított ingatlanok aránya (42,0%) valamelyest magasabb a teljes mintában megjelenő arányukhoz (39,1%) viszonyítva,
- az 55-59 dB-es zajzónán belül a felújítást igénylő ingatlanok aránya (20,3%) valamelyest magasabb a teljes mintában megjelenő arányukhoz (17,1%) viszonyítva,
- és a 60 dB zajzónában lévő ingatlanok között tipikusnak tekinthetők az új építésű ingatlanok: az új építésű ingatlanok aránya (47,2%) kétszer akkora a 60 dB feletti zajzónán belül, mint a teljes mintán belül (23,6%).

E) Készítsen grafikus ábrát!

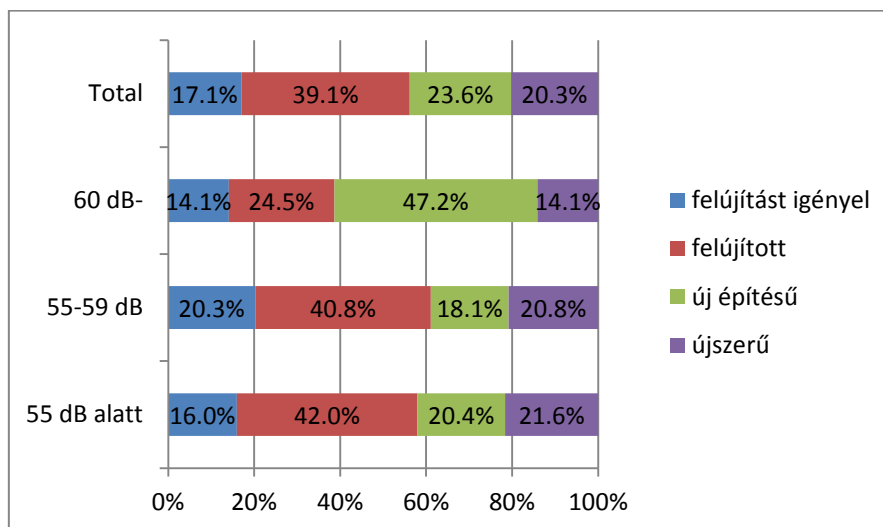
Az arányok megjelenítéséhez célszerű az Excel 100%-ig halmozott szalagdiagramját alkalmazni. Ehhez a megoszlásokat tartalmazó táblát kell átmásolni Excelbe.

Állapot * Zajszint (dB) Crosstabulation

% within Zajszint (dB)

		Zajszint (dB)			Total
		55 dB alatt	55-59 dB	60 dB-	
Állapot	felújítást igényel	16,0%	20,3%	14,1%	17,1%
	felújított	42,0%	40,8%	24,5%	39,1%
	új építésű	20,4%	18,1%	47,2%	23,6%
	újszerű	21,6%	20,8%	14,1%	20,3%
Total		100,0%	100,0%	100,0%	100,0%

Amelyre a 100%-ig halmozott sáv diagramtípus illesztésével az alábbi ábra jön létre:



2. feladat

A repter.xls állomány egy repülőtér közelében eladásra kínált ingatlanok véletlen, reprezentatív mintáját tartalmazza. A környék zajterhelésének és az ingatlan állapotának kapcsolatát vizsgálva válaszoljon Excel segítségével az alábbi kérdésekre.

A 2. feladat végeredményei az 1. feladat végeredményeivel megegyeznek, a szöveges értelmezés csak az 1. feladatnál kerül feltüntetésre. A megoldások között az Excel-lel létrehozott eredmények képernyőrészletei találhatóak meg.

A) Írassa ki zajszintet és az állapotot tartalmazó kombinációs táblát!

Mennyiség / állapot	zajszint			Végösszeg
	55 dB alatt	55-59 dB	60 dB-	
felújítást igényel	93	74	23	190
felújított	245	149	40	434
új építésű	119	66	77	262
újszerű	126	76	23	225
Végösszeg	583	365	163	1111

B) Írassa ki az állapot szerinti megoszlást a különböző zajzónákban külön-külön!

Mennyiség / állapot	zajszint			Végösszeg
	55 dB alatt	55-59 dB	60 dB-	
felújítást igényel	15,95%	20,27%	14,11%	17,10%
felújított	42,02%	40,82%	24,54%	39,06%
új építésű	20,41%	18,08%	47,24%	23,58%
újszerű	21,61%	20,82%	14,11%	20,25%
Végösszeg	100,00%	100,00%	100,00%	100,00%

C) Számítsa ki a függetlenség esetén fennálló gyakoriságokat és a kapcsolatmérő mutatót (Cramer-együttható)!

A függetlenség esetén fennálló gyakoriságok kiszámításához először a gyakoriságokat tartalmazó kombinációs táblára lesz szükség, amelyből a gyakoriságokat felhasználva az alábbi módon határozhatók meg a függetlenség esetén fennálló gyakoriságok:

3	Mennyiség / állapot	zajsztint			
4	állapot	55 dB alatt	55-59 dB	60 dB-	Végösszeg
5	felújítást igényel	93	74	23	190
6	felújított	245	149	40	434
7	új építésű	119	66	77	262
8	újszerű	126	76	23	225
9	Végösszeg	583	365	163	1111
10					
11					
12					
13	Függetlenség esetén fennálló gyakoriságok:				
14					
15	állapot	55 dB alatt	55-59 dB	60 dB-	Összesen
16	felújítást igényel	=B\$9*\$E5/\$E\$9			
17	felújított				
18	új építésű				
19	újszerű				
20	Összesen				

Megjegyzés:

- a kimutatás esetén a cellák kijelölését célszerű nem egérrel való kattintással végrehajtani, hanem begépelni (=B\$9*\$E5/\$E\$9)
- Ahhoz, hogy az Excelben a számítás esetén az automatikus kitöltést (a képlet „lefele és jobbra húzását”) tudjuk alkalmazni, az oszlopok peremgyakoriságai (9-es sor) esetén a sor számát kell rögzíteni, míg a sorok peremgyakoriságai (E oszlop) esetén az oszlopot kell rögzíteni, valamint a minta elemszáma esetén a teljes cellát (E9-es cella) rögzíteni kell. Végeredményként az alábbi függetlenség esetén fennálló gyakoriságokat kapjuk:

állapot	55 dB alatt	55-59 dB	60 dB-	Összesen
felújítást igényel	99,7	62,4	27,9	190
felújított	227,7	142,6	63,7	434
új építésű	137,5	86,1	38,4	262
újszerű	118,1	73,9	33,0	225
Összesen	583	365	163	1111

- Mivel a Cramer-együttható meghatározásához és a D) kérdés megválaszolásához is szükség van a Khí-négyzet függvény kiszámítására, így a Cramer-együttható kiszámítását is a D) pont keretében mutatjuk be.

D) Ötszázalékos szignifikanciaszint mellett szignifikáns kapcsolat van-e a zajsztint és az állapot között?

Khí-négyzet próbafüggvény mintán felvett értéke	63,327
Empirikus szignifikancia szint	9,46617E-12
Cramer-együttható	0,169

Megjegyzés:

- A Khí-négyzet teszt végrehajtása nincs beépítve az Excel Adatelemzés bővítménybe. E-helyett nekünk kellene kézzel kiszámítanunk a Khí-négyzet próbafüggvény mintán felvett

értékét. A függetlenségvizsgálat esetén használt próbafüggvény az alábbi formára hozható:

$$\begin{aligned} \chi^2 &= \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*} = \sum_{i=1}^r \sum_{j=1}^c \frac{f_{ij}^2 - 2 \cdot f_{ij} \cdot f_{ij}^* + f_{ij}^{*2}}{f_{ij}^*} = \\ &= \sum_{i=1}^r \sum_{j=1}^c f_{ij} \frac{f_{ij}}{f_{ij}^*} - 2 \sum_{i=1}^r \sum_{j=1}^c f_{ij} + \sum_{i=1}^r \sum_{j=1}^c f_{ij}^* = \sum_{i=1}^r \sum_{j=1}^c f_{ij} \frac{f_{ij}}{f_{ij}^*} - 2 \cdot n + n = \\ &= \sum_{i=1}^r \sum_{j=1}^c f_{ij} \cdot \frac{f_{ij}}{f_{ij}^*} - n \end{aligned}$$

KIMUTATÁS INDEXMEGJELENÍTÉS

Az utolsó tag azt jelenti, hogy a próbafüggvény mintán felvett értékét ki tudjuk számítani az Excel kimutatás készítésével (az Index megjelenítési mód alkalmazásával). A függvényben szereplő első tag gyakorlatilag a gyakoriságok és az *Index* értékek szorzatösszege, melyet az Excelben a *szorzatösszeg* függvény segítségével tudunk kiszámítani, majd a szorzatösszegeből kell kivonni a minta elemszámát.

2					
3	Mennyiség / állapot	zajsztint			
4	állapot	55 dB alatt	55-59 dB	60 dB-	Végösszeg
5	felújítást igényel	0,932770606	1,185493872	0,825088796	1
6	felújított	1,075776019	1,045003472	0,628198241	1
7	új építésű	0,865548034	0,766767751	2,003161148	1
8	újszerű	1,067169811	1,02814003	0,69674165	1
9	Végösszeg	1	1	1	1
10					
11					
12	állapot	55 dB alatt	55-59 dB	60 dB-	Végösszeg
13	felújítást igényel	93	74	23	190
14	felújított	245	149	40	434
15	új építésű	119	66	77	262
16	újszerű	126	76	23	225
17	Végösszeg	583	365	163	1111
18					
19	Khí-négyzet próbafüggvény mintán felvett értéke	=SZORZATÖSSZEG(B5:D8;B13:D16)-E17			
20	Empirikus szignifikancia szint				
21	Cramer-együttható				
22					

- Az empirikus szignifikancia értéket a *Khí.eloszlás(a próbafüggvény mintán felvett értéke, szabadságfok)* függvény segítségével számíthatjuk ki, ahol a szabadságfok nem más, mint $(a \text{ kombinációs tábla adatsorainak száma}-1) * (a \text{ kombinációs tábla adatoszlopainak száma}-1)$, azaz $(r-1)*(c-1)$.

2					
3	Mennyiség / állapot	zajszint			
4	állapot	55 dB alatt	55-59 dB	60 dB-	Végösszeg
5	felújítást igényel	0,932770606	1,185493872	0,825088796	1
6	felújított	1,075776019	1,045003472	0,628198241	1
7	új építésű	0,865548034	0,766767751	2,003161148	1
8	újszerű	1,067169811	1,02814003	0,69674165	1
9	Végösszeg	1	1	1	1
10					
11					
12	állapot	55 dB alatt	55-59 dB	60 dB-	Végösszeg
13	felújítást igényel	93	74	23	190
14	felújított	245	149	40	434
15	új építésű	119	66	77	262
16	újszerű	126	76	23	225
17	Végösszeg	583	365	163	1111
18					
19	Khí-négyzet próbafüggvény mintán felvett értéke	63,327			
20	Empirikus szignifikancia szint	=KHI.ELOSZLAS(B19;6)			
21	Cramer-együttható				

Az Excel az empirikus szignifikancia szintet jelen esetben normálalakban közölte, tehát a megjelenített érték: $9,46 \cdot 10^{-12}$. Amennyiben az SPSS kimenetre duplán kattintunk, majd azon belül a Pearson Chi-Square sor sig értékére is duplán kattintunk, ugyanazt az értéket fogjuk látni, mint ami az Excelben megjelent.

- A Cramer-együtthatót pedig a GYÖK függvény segítségével határozhatjuk meg az alábbi módon:

$$C = \sqrt{\frac{63,327}{1111 \cdot \min\{(4-1), (3-1)\}}} = \sqrt{\frac{63,327}{1111 \cdot 2}} = 0,169$$

11					
12	állapot	55 dB alatt	55-59 dB	60 dB-	Végösszeg
13	felújítást igényel	93	74	23	190
14	felújított	245	149	40	434
15	új építésű	119	66	77	262
16	újszerű	126	76	23	225
17	Végösszeg	583	365	163	1111
18					
19	Khí-négyzet próbafüggvény mintán felvett értéke	63,327			
20	Empirikus szignifikancia szint	9,46617E-12			
21	Cramer-együttható	=GYÖK(B19/(E17*2))			
22					

3. feladat

A bank.sav állomány egy vállalat dolgozóinak véletlen mintáját tartalmazza. Válaszoljon SPSS segítségével az alábbi kérdésekre.

A) Írassa ki az alkalmazottak foglalkozás szerinti megoszlását nemenként külön-külön!

foglalkozás * nem Crosstabulation

		nem		Total	
		férfi	nő		
foglalkozás	alapfokú végzettségű beosztott	Count	27	0	27
		% within nem	10,5%	0,0%	5,7%
	középfokú végzettségű beosztott	Count	157	206	363
		% within nem	60,9%	95,4%	76,6%
	felsőfokú végzettségű beosztott	Count	64	9	73
		% within nem	24,8%	4,2%	15,4%
	középfokú végzettségű vezető	Count	4	1	5
		% within nem	1,6%	0,5%	1,1%
	felsőfokú végzettségű vezető	Count	6	0	6
		% within nem	2,3%	0,0%	1,3%
Total	Count	258	216	474	
	% within nem	100,0%	100,0%	100,0%	

B) Ötszázalékos szignifikanciaszint mellett szignifikáns kapcsolat van-e a nem és a foglalkozás között? Amennyiben igen vizsgálja meg valamely kapcsolatmutató értékét is!

foglalkozás * nem Crosstabulation

		nem		Total	
		férfi	nő		
foglalkozás	alapfokú végzettségű beosztott	Count	27	0	27
		Expected Count	14,7	12,3	27,0
		% within nem	10,5%	0,0%	5,7%
	középfokú végzettségű beosztott	Count	157	206	363
		Expected Count	197,6	165,4	363,0
		% within nem	60,9%	95,4%	76,6%
	felsőfokú végzettségű beosztott	Count	64	9	73
		Expected Count	39,7	33,3	73,0
		% within nem	24,8%	4,2%	15,4%
	középfokú végzettségű vezető	Count	4	1	5
		Expected Count	2,7	2,3	5,0
		% within nem	1,6%	0,5%	1,1%
	felsőfokú végzettségű vezető	Count	6	0	6
		Expected Count	3,3	2,7	6,0
		% within nem	2,3%	0,0%	1,3%
Total	Count	258	216	474	
	Expected Count	258,0	216,0	474,0	
	% within nem	100,0%	100,0%	100,0%	

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	79,757 ^a	4	,000
Likelihood Ratio	97,263	4	,000
Linear-by-Linear Association	12,786	1	,000
N of Valid Cases	474		

a. 4 cells (40,0%) have expected count less than 5. The minimum expected count is 2,28.

Mivel a cellák 40 százalékában (lásd táblázat alatti megjegyzés) a független gyakoriságok értéke 5-nél alacsonyabb, így a teszt alkalmazása nem teljesül, ezért kategóriákat (a középfokú végzettségű és a felsőfokú végzettségű vezetők csoportjait) szükséges összevonni. Az összevontást újrakódolással hajthatjuk végre (SPSS: recode into same variable vagy recode into different variable, Excel: automatikus csere az adott oszlopban). Az újrakódolás után a teszt alkalmazási feltétele teljesült.

foglalkozás * nem Crosstabulation

		nem		Total	
		férfi	nő		
foglalkozás	alapfokú végzettségű beosztott	Count	27	0	27
		Expected Count	14,7	12,3	27,0
		% within nem	10,5%	0,0%	5,7%
	középfokú végzettségű beosztott	Count	157	206	363
		Expected Count	197,6	165,4	363,0
		% within nem	60,9%	95,4%	76,6%
	felsőfokú végzettségű beosztott	Count	64	9	73
		Expected Count	39,7	33,3	73,0
		% within nem	24,8%	4,2%	15,4%
	vezető	Count	10	1	11
		Expected Count	6,0	5,0	11,0
		% within nem	3,9%	0,5%	2,3%
Total	Count	258	216	474	
	Expected Count	258,0	216,0	474,0	
	% within nem	100,0%	100,0%	100,0%	

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	79,318 ^a	3	,000
Likelihood Ratio	95,565	3	,000
Linear-by-Linear Association	11,972	1	,001
N of Valid Cases	474		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,01.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,409	,000
	Cramer's V	,409	,000
N of Valid Cases		474	

Az alkalmazási feltételt kielégítő megoldásra az **alábbi összefüggő szöveges elemzést** fogalmazhatjuk meg:

Mivel két kategoriális változó közötti kapcsolatot vizsgálunk, ezért kereszttábla elemzést használhatunk a probléma vizsgálatára. A vizsgálat nullhipotézise szerint a foglalkozás és a nem között nincs szignifikáns kapcsolat. A próbafüggvény alkalmazási feltétele a cellák több mint 20%-nál nem teljesült, így kategóriákat kellett összevonni, azaz a közép és a felsőfokú végzettségű vezetők csoportja összevonásra került. Ezután a teszt alkalmazási feltétele teljesült.

A nem és a foglalkozás között ötszázalékos szignifikanciaszinten kimutatható szignifikáns ($\text{sig} < 0,05$, $\chi^2 = 79,318$), erős ($C = 0,409$) kapcsolat. A férfiak között a felsőfokú végzettségű beosztottak csoportja tekinthető inkább tipikus kategóriának, mivel a férfiak között felsőfokú végzettségű beosztottak aránya 24,8% (míg a teljes mintában ugyanez az arányszám 14,5%). A nők között pedig a középfokú végzettségű beosztottak csoportja tekinthető inkább tipikus kategóriának, mivel a nők között a középfokú végzettségű beosztottak aránya 95,4% (míg ugyanez az arányszám a teljes mintában 76,6%).

SZEGEDI TUDOMÁNYEGYETEM
GAZDASÁGTUDOMÁNYI KAR
KÖZGAZDÁSZ KÉPZÉS
TÁVOKTATÁSI TAGOZAT
LECKESOROZAT
COPYRIGHT © SZTE GTK 2017/2018

A LECKE TARTALMA, ILLETVE ALKOTÓ ELEMEI ELŐZETES,
ÍRÁSBELI ENGEDÉLY MELLETT HASZNÁLHATÓK FEL.

JELLEN TARTALOM
A SZEGEDI TUDOMÁNYEGYETEMEN KÉSZÜLT
AZ EURÓPAI UNIÓ TÁMOGATÁSÁVAL.
PROJEKT AZONOSÍTÓ: EFOP-3.4.3-16-2016-00014

SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE