

Név:

Neptun kód:

Töltse ki ezt a feladatlapot, és mentse el a fájlt. A fájl neve kezdődjön a saját nevével. Minden munka legyen ebben az egy fájlban. A kitöltött dolgot ezen a linken lehet feltölteni:

<https://www.dropbox.com/request/sJyoHITPk3oZbVBBopi>

1. feladatcsoport: : Darts (35/100 pont)

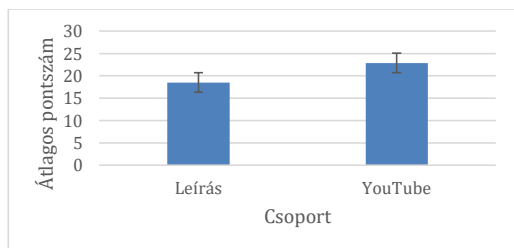
Adatfile: [Darts.csv](#)

Kardas & O'Brien (2018) arra volt kíváncsi, hogy tudunk-e ügyességi feladatokat Youtube videókból tanulni. Két csoportnak darts (nyíldobálós) játékot tanítottak. Az egyik csoport többször megnézett egy Youtube videót, amin a játékot bemutatták, a másik csoport pedig egy leírást olvasott a játék menetéről. A tanítás után a kísérleti személyeknek próbára kellett tenniük az ügyességüket a játékban, és a függő változó az elért pontszám volt.

1.1 Fogalmazzon meg egy hipotézist. (1 pont)

A Youtube videóból tanulók jobban teljesítenek, mint a leírásból tanulók. VAGY A leírásból tanulók jobban teljesítenek, mint a Youtube videóból tanulók. VAGY Különbség lesz a Youtube videóból és a leírásból tanulók darts pontszámai között.

1.2 Rajzoljon a hipotézisnek megfelelő ábrát. (9 pont)



1.3 Ellenőrizze az eloszlást, számolja ki (olvassa le a JASP-ból) a megfelelő konfidencia intervallumokat és hatásméretet, és írja le röviden az eredményeket. (25 pont)

Commented [BA1]: oszlopdiagram átlagokkal és hiba barokkal (szórás, standard hiba vagy konfidencia intervallum). Tengelyek felcímkézve.

A kísérlet azt vizsgálta, hogy videóból jobban lehet-e tanulni ügyességi játékot, mint leírásból. Két csoportnak darts (nyíldobálós) játékot tanítottak. Az egyik csoport többször megnézett egy Youtube videót (N = 29), amin a játékot bemutatták, a másik csoport pedig egy leírást olvasott a játék menetéről (N = 29). A tanulás után lemérték mindkét csoport készségét.

A csoportok pontszámai normál eloszlást mutatnak. A leírásból tanuló csoport (M = 18,52; CI: 15,17; 21,86) átlagosan kevesebb pontot ért el, mint a videóból tanuló csoport (M = 22,90; CI: 19,73; 26,07). Az átlagok konfidencia intervallumának jelentős átfedése azonban arra utal, hogy nincs általánosítható különbség a két csoport között. Ezt a konklúziót erősíti meg a hatásméret konfidencia intervalluma is (d = - 0,49; CI: -1,01; 0,03), ami azt mutatja, hogy a populációra vonatkoztatva akár az is lehetséges, hogy a leírásból tanulók teljesítenek jobban. Az adataink tehát nem erősítik meg a hipotézist miszerint videóból jobb ügyességi játékot tanulni, mint leírásból.

Commented [BA2]: Hipotézis

Commented [BA3]: Adatgyűjtés módszere, független változó szintjei, minták elemszáma

Commented [BA4]: Fügő változó definiálása

Commented [BA5]: Átlagok és konfidencia intervallumok jelentése

Commented [BA6]: Átlag konfidencia intervallumának értelmezése

Commented [BA7]: Hatásméret és értelmezése

Commented [BA8]: Konklúzió, visszautalás a hipotézisre

2. feladatcsoport: Jövedelem (65/100 pont)

Adatfile: [GSS2018Income.csv](#)

A fájlban az 2018-as General Social Survey adataiból látszik néhány változó: jövedelem (szterdendizált dollárban), az elvégzett iskolai évek száma, az állásban eltöltött évek száma, a gyenge mentális egészségben töltött napok száma az elmúlt hónapban, és hogy egy héten hány órát néz TV-t a válaszadó.

2.1 Vizsgálja meg, hogy melyik változók korrelálnak a jövedelemmel, és ezek alapján fogalmazzon meg egy sor hipotézist a változók jövedelemre tett hatására vonatkozóan. (20 pont)

A változók mind skála változók, de egyik sem normál eloszlású. A jövedelem változó sok személynél ugyanazt az erősen kiugró értéket veszi fel, ami mindenképpen gyanús. Ezeket kizárom a további elemzésből. A hipotézisek felállításához Spearman korrelációelemzéseket végzek. Ezek eredménye szerint az állásban eltöltött idő korrelál a legerősebben a jövedelemmel, ezt követi az oktatásban töltött évek száma. A TV-nézés gyengécske, nem szignifikáns negatív korrelációt mutat, a mentális problémák gyakorisága teljesen függetlennek tűnik.

Ezek alapján a hipotéziseim: A jövedelmet pozitívan befolyásolja elsősorban az állásban töltött évek száma és másodsorban az iskolázottság. A tévénézésnek lehet esetleg negatív hatása.

2.2 Építsen regressziós modellt a hipotézisek tesztelésére, és foglalja össze a főbb eredményeket egy táblázatban. (10 pont)

Két modellem lesz: az elsőben (a null modellben) az állásban töltött évek száma, a másodikban pedig mindhárom változó (állásban évek, iskolázottság, tévénézés).

	b	Beta	t	p
1. modell				
Állás-évek	252,57	0,14	3,91	<,001
2. modell				
Állás-évek	227,24	0,12	3,53	<,001
Iskolázottság	849,86	0,14	4,01	<,001
Tévénézés	-171,74	-0,02	-0,59	,56

2.3 Értelmezze az eredményeket, és írja le röviden, a tudományos közlemények formátumában, hogy milyen következtetéseket tehetünk. (35 pont)

Commented [BA9]: Ha valakinek itt nem tűnt fel, hogy probléma van, a regresszióelemzésnél fel kell tűnnie: ha ezek az esetek maradnak, a reziduumok eloszlása nagyon nem jó, akár a hisztogram, akár a Casewise diagnostics kiadja ezeket a kiugró értékeket. Ezen kívül a homoszkedaszticitás is sérül miattuk (Residuals vs. predicted ábra). Miután kizártuk ezeket a kiugró értékeket, minden rendben lesz.

Commented [BA10]: Más hipotézis is lehetséges, de valamilyen indoklás kell hozzá.

Commented [BA11]: Az is jó megoldás, ha valaki az iskolázottságot teszi az első modellbe, mert azt tartja a legfontosabb tényezőnek. Lehet ezen kívül a JASP-ra bízni a modellépítést (forward vagy backward módszerrel). Az viszont nem túl jó megoldás, ha csak bedobjuk mind a négy változót egy modellbe.

Commented [BA12]: az interceptet is be lehet tenni a táblázatba, bár itt valószínűleg nem sok értelme van.

A 2018-as General Social Survey adatai alapján azt vizsgáltuk lineáris regresszióelemzéssel, hogy milyen tényezők befolyásolják egy egyén jövedelmét. A jövedelmet sztenderdizált dollárban nézzük. A magyarázó változóink az elvégzett iskolai évek száma, az állásban eltöltött évek száma, és hogy egy héten hány órát néz TV-t a válaszadó. A felmérésben 810 személynek szerepel valamennyi ide vonatkozó adata. Ebből 29 személy jövedelme több mint három szórással magasabb az átlagnál, ezért őket kizárjuk a további elemzésekből. Így 781 válaszadó maradt a mintában. Előzetes vizsgálódásaink alapján az állásban töltött idő tűnik a leginkább meghatározó tényezőnek, ezért az első modellbe csak ezt a változót léptettük be. A második modellben mindhárom magyarázó változó szerepel.

Mindkét modell szignifikánsnak bizonyult (1. modell: $F(1, 779) = 15,27$; $p < ,001$; $R^2 = 0,02$; 2. modell: $F(1, 777) = 10,87$; $p < ,001$; $R^2 = 0,04$). A 2. modellből kiderült, hogy a tévénézés nincs szignifikáns hatással a jövedelemre, az állásban eltöltött idő és az iskolázottság viszont szignifikánsan, pozitívan befolyásolja. Az 1. modell a jövedelem varianciájának 2 százalékát, a 2. modell pedig 4 százalékát magyarázza. A korrelációelemzés eredményével ellentétben a regressziós modellek arra utalnak, hogy iskolai évek száma a legjelentősebb tényező: minden szórányi növekedés az oktatásban töltött évek számában 0,14 szórással növeli a jövedelmet. Az állásban eltöltött évek számának ennél valamivel gyengébb a hatása: ez a tényező 0,12 szórással növeli a jövedelmet. Összességében tehát azt a konklúziót vonhatjuk le, hogy egyrészt érdemes az oktatásba befektetni, másrészt pedig érdemes egy munkahelyen hosszabb távon is megmaradni, bár a két változó együtt is csak viszonylag kis mértékben járul hozzá a jövedelem varianciájának magyarázatához. Mivel az eredményeink amerikai adatokra épülnek, ezek a konklúziók nem feltétlenül érvényesek a világ más tájaira.

Commented [BA13]: téma, adatok eredete

Commented [BA14]: függő változó

Commented [BA15]: magyarázó változók felsorolása

Commented [BA16]: elemszám, kizárások

Commented [BA17]: hipotézis, modellépítés menete

Commented [BA18]: modellek szignifikanciája, R négyzet

Commented [BA19]: melyik változók hatása szignifikáns, és melyeké nem az

Commented [BA20]: R négyzet értelmezése

Commented [BA21]: együthetők értelmezése

Commented [BA22]: Kolinearitás és Durbin-Watson rendben van, ezért meg sem kell említeni. A reziduumok eloszlása és homoszkedaszticitás is rendben van, ha a kiugró értékek ki vannak zárva. Ha nem sikerült őket kizárni az előző lépésben, akkor most a Casewise diagnostics és a Residuals vs. predicted ábra kimutatja őket. A helyes eljárás ezesetben az, hogy kizárjuk őket és kezdjük elől a regressziós modell építését.

Commented [BA23]: konklúzió, visszautalás a hipotézisekre