Mariokart

Miért mindenki Waluigival játszik online?

Többváltozós adatelemzési modellek (MAME039LMSB) házidolgozat

Fazekas Márk Máté

EUV0J3

**Tartalomjegyzék**

[Az adatbázis bemutatása 2](#_Toc125223518)

[Leíró statisztikai elemzés 3](#_Toc125223519)

[Intervallumbecslés és Hipotézisvizsgálat 3](#_Toc125223520)

[Kétváltozós kapcsolatvizsgálat 4](#_Toc125223521)

**A dolgozat formai követelményei:**

* Ügyeljünk a formaiság minimális szintjére
* exportált táblázatokkal
* grafikonokkal illusztrálva)
* Az ábrák, táblázatok igényesen legyenek elhelyezve!

# Az adatbázis bemutatása

Forrás: <https://www.kaggle.com/datasets/marlowspringmeier/mario-kart-8-deluxe-ingame-statistics> (Letöltés dátuma: 2023.01.21)

Röviden mutassa be, hogy mik az adatbázis megfigyelési egységei (sorai) és adja meg az adatbázis változóinak (oszlopainak) pontos jelentését és mérési skáláját (nominális, ordinális, intervallum, arány)! Numerikus változók esetén mindenképpen adja meg az

adatok mértékegységét is!

minimum 6 változót (más néven ismérvet)

legalább 50 megfigyelési egységet tartalmaz

A változók közül minimum 2 legyen minőségi (más néven nominális vagy „szöveges”)

Legalább egy minőségi változónak legyen minimum 3 lehetséges értéke

A többi változó közül legalább 3 legyen numerikus!

# Leíró statisztikai elemzés

Az adatbázisban lévő változók közül válasszon ki egy numerikusat, és jellemezze a változót a következő statisztikai eszközökkel: gyakorisági táblázat (hisztogram is), helyzetmutatók, kvantilisek (doboz ábra is), szóródási mutatók, alakmutatók!

A külső és belső kerítések segítségével keressen kiugró értékeket a kiválasztott mennyiségi ismérvek esetében!

Minden statisztikai elemzés eredményét értelmezze is, és vesse is össze az eredményeket egymáshoz képest (pl. mit árul az átlag-medián viszonya és hogyan olvasható le ez a hisztogramról)!

Válasszon ki a változók közül egy nem numerikusat, és adja meg, hogy mely leíró statisztikai módszerek értelmezhetőek a változó mérési skáláján! A módszereket alkalmazza a kiválasztott változón, és értelmezze az eredményüket!

# Intervallumbecslés és Hipotézisvizsgálat

Válasszon ki egy numerikus és minőségi változót és végezze el a numerikus változó átlagának és mediánjának intervallumbecslését legalább 90%-os megbízhatósági szinten a minőségi változó kategóriái szerint külön-külön!

oA konfidencia-intervallumokat értelmezze, és az eredmények összehasonlítása segítségével vonjon le egyszerű következtetéseket!

Válasszon ki egy minőségi változót és azon belül egy konkrét kategóriát. Végezzen a kategória arányára legalább 90%-os megbízhatóságú intervallumbecslést egy másik, szintén minőségi változó kategóriái szerint külön-külön!

oA konfidencia-intervallumokat értelmezze, és az eredmények összehasonlítása segítségével vonjon le egyszerű következtetéseket!

Vizsgálja meg, hogy az intervallumbecslések elvégzéséhez szükséges előfeltételek teljesülnek-e (megengedett, hogy azok ne teljesüljenek)!

Adjon meg egy olyan hipotézist (állítást), ami a választott adatbázison elvégzett egymintás átlagra, arányra vagy mediánra vonatkozó próbák segítségével vizsgálható!

Adja meg, hogy a nullhipotézis elutasítása vagy elfogadása mellett elfogadható az

eredetileg megfogalmazott állítás!

Adja meg a próbákhoz tartozó p-értékeket, és döntsön az eredeti hipotézisről a szokásos

szignifikancia-szintek mellett! Értékelje a döntések stabilitását!

Vizsgálja meg, hogy a hipotézisvizsgálatok elvégzéséhez szükséges előfeltételek

teljesülnek-e (megengedett, hogy azok ne teljesüljenek)!

# Kétváltozós kapcsolatvizsgálat

Az adatbázisban lévő változók közül válasszon ki két minőségi változót, majd elemezze azok kapcsolatát a következők szerint:

írja le a kapcsolat jellegét a megfelelő diagram segítségével!

minősítse a kapcsolat szorosságát a mintán belül Cramer együttható

segítségével!

a tanult módon tesztelje a kapcsolat fennállását a sokaságban!

ellenőrizze a próba előfeltételeit (megengedett, hogy azok ne teljesüljenek)! értelmezze a két eredményt külön-külön és együtt is!

Az adatbázisban lévő változók közül válasszon ki egy mennyiségi és egy minőségi változót, majd elemezze azok kapcsolatát a következők szerint:

írja le a kapcsolat jellegét a megfelelő diagram segítségével!

adjamegésértelmezzeazadatbázisszövegkörnyezetébenakülső,belsőésteljes

szórások értékét!

minősítse a kapcsolat szorosságát a mintán belül a variancia-hányados

segítségével!

a tanult módon tesztelje a kapcsolat fennállását a sokaságban!

ellenőrizze a próba azon előfeltételét, amelyet a tanult eszközökkel el tud

végezni (megengedett, hogy azok ne teljesüljenek)! értelmezze a két eredményt külön-külön és együtt is!

Az adatbázisban lévő változók közül válasszon ki két numerikus változót, majd elemezze azok kapcsolatát a következők szerint:

írja le a kapcsolat jellegét a megfelelő diagram segítségével!

minősítse a kapcsolat szorosságát és irányát a mintán belül a korrelációs és a

determinációs együttható segítségével!

határozza meg és értelmezze a regressziós egyenesek paramétereit!

minősítse a regressziós egyenes becslési pontosságát a reziduális standard hiba

segítségével!

a tanult módon tesztelje a kapcsolat fennállását a sokaságban!

vonjon le az eredményekből egyszerű következtetéseket!

a két kiválasztott numerikus változó mellé válasszon egy harmadik, minőségi

változót, és vizsgálja meg grafikusan és hipotézisvizsgálat segítségével is, hogy a két numerikus változó kapcsolata eltérően viselkedik-e a minőségi változó csoportjaiban!