Mario Kart

Többváltozós adatelemzési modellek (MAME039LMSB) házidolgozat

Fazekas Márk Máté

EUV0J3

**Tartalomjegyzék**

[Bevezetés 2](#_Toc125392596)

[Az adatbázis bemutatása 3](#_Toc125392597)

[Score elemzése (Leíró statisztikai elemzés) 5](#_Toc125392598)

[Size (nem numerikus) 5](#_Toc125392599)

[Intervallumbecslés és Hipotézisvizsgálat 6](#_Toc125392600)

[Kétváltozós kapcsolatvizsgálat 6](#_Toc125392601)

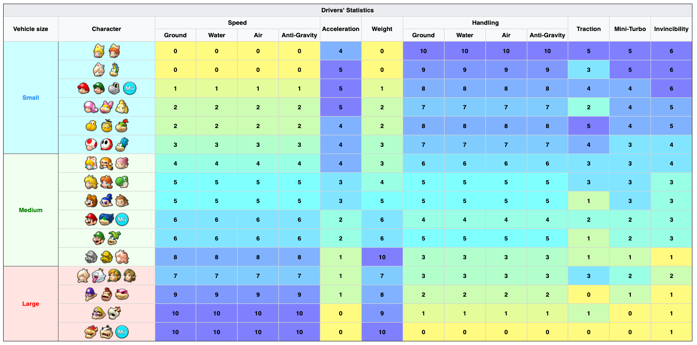
# Bevezetés

A Mario Kart egy Nintendo által fejlesztett és forgalmazott versenyzős játék.A Mario Kart 8 eredeti kiadását 2014-ben hozták forgalomba, a Delux változat (ami az előző frissítése), 2017-ben került a nyilvánosság elé. A mai napig frissíti a Nintendo, ezzel a felhasználókat visszacsábítva időről időre. Ez a játék nem csak az átlagos felhasználók, de az úgynevezett „speed-runnerek” által is közkedvelt. Ezek olyan felhasználók, akik minél jobb időt szeretnének elérni, a bizonyos versenypályákon.

Egy versenyző eredményét a tradicionális futamokon több dolog is befolyásolhatja, de a versenyek a minél jobb időkért általában az időmérő futamokon szokott történni. Itt tényleg csak a versenyzőn múlik minden. Az egyik fontos döntés, amit a versenyző meghozhat, hogy milyen „konfigurációval” fog versenyezni.

Minden versenyzőnek választania kell egy sofőrt (driver), egy járművet (kart), egy fajta kereket (tire) és egy ernyőt (glider). Ennek a 4 komponens összetételével különböző járművek kreálhatóak, amik a különböző tulajdonságaik miatt, különböző eredményekhez vezethetnek. A játék ezeket a tulajdonságokat számokkal jellemzi, és grafikonokon vizualizálja a játékos számára. A továbbiakban azt fogom vizsgálni, hogy átlagos játékosoknak, milyen kombináció lehet a leghatékonyabb. Mivel több felületen folyik a verseny a játékban (szárazföld, víz, levegő, mesterséges gravitáció), ezért a pontos elemzéshez a pályák pontos összetétele is szükséges lenne, ami jelenleg nem elérhető, így az egyszerűség kedvéért, csak a szárazföldi paramétereket elemzem.

# Az adatbázis bemutatása

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

A kaggle-n elérhető adatbázissal könnyebb volt elkezdeni dolgozni, de a MarioWikin, gyakrabban frissített adatok érhetőek el.

Az eredeti adatbázis négy táblából áll (driver, kart, tire, glider). A négy táblának Descartes-szorzata és oszlopainak egyszerűsítésével generáltam az elemzett adatbázist.

ábra 1 Sofőrök statisztikái (MarioWiki)

Egy megfigyelési egység egy lehetséges kombináció, amit a játékos választhat (43\*40\*21\*14 = 505 680 megfigyelés):

* Driver
  + A választott karakter (pl.: Mario)
  + 43 különböző érték
* Kart
  + A választott jármű (pl.: Standard Kart)
  + 40 különböző érték
* Tire
  + A választott kerék (pl.: Standard)
  + 21 különböző érték
* Glider
  + A választott ernyő (pl.: Super Glider)
  + 14 különböző érték

És a hozzá tartozó mutatók:

* Size
  + A driver mérete
  + Ordinális minőségi változó
    - Lehetséges értékek: small / medium / large
* Weight
  + A konfiguráció súlya. A nehezebb konfigurációkkal, ha kiütünk egy játékost, tovább tart az ellenfélnek újra indulnia. A nehezebb játékosokat mikor kiütik, ők is nehezebben indulnak újra.
  + Különbségi (intervallum) mérési skála
* Speed
  + A konfiguráció maximum sebessége.
  + Különbségi (intervallum) mérési skála
* Acceleration
  + A konfiguráció gyorsulása. A maximum sebesség növekedés képkockánként
  + Különbségi (intervallum) mérési skála
* Handling
  + A konfiguráció sebessége kanyarodás közben.
  + Különbségi (intervallum) mérési skála
* Traction
  + A konfiguráció irányíthatósága
  + Különbségi (intervallum) mérési skála
* Score
  + Speed + Acceleration + Handling + Traction
  + Különbségi (intervallum) mérési skála

Graphical user interface, application, table, Excel

Description automatically generated

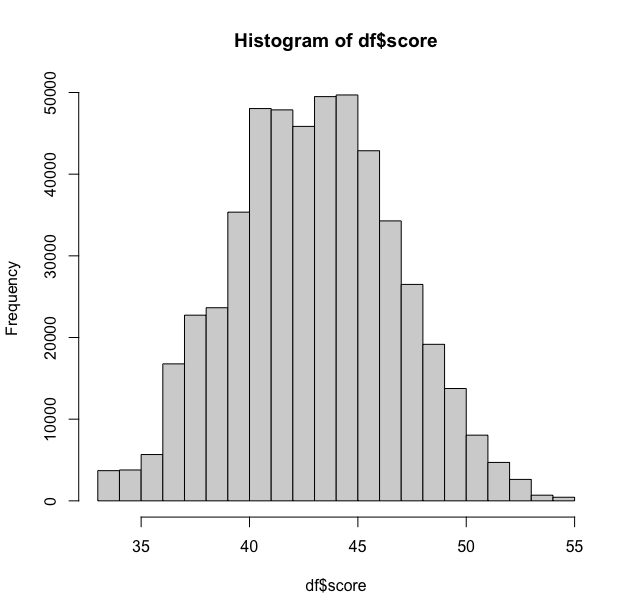
ábra 2 df

# Score elemzése (Leíró statisztikai elemzés)

A score 33 és 55 közötti értékek lehetnek. Az elemszám itt is 505 680, mivel minden konfigurációnak van pontja.

A hisztogramról és a gyakorisági ábráról látható, hogy a leggyakoribb érték a 45 (módusz).

A medián 43 (azaz a konfigurációk fele ennél rosszabb és a másik fele ennél jobb) és az átlag 43.48. Azaz, egy átlagos konfiguráció 43.48 pontos.

A hisztogramról és az átlag, módusz és a medián viszonyából feltételezem, hogy az adat szimmetrikus eloszlású. Az α3 (skew) értéke 0.04, ez igazolja a feltételezést, hogy szimmetrikus eloszlású. Ez ugye azt is jelenti, hogy a változó értékei véletlen hatások összegződésével állnak elő.

Az α4 (kurtosis) értéke -0.32, ami azt jelenti, hogy a változó eloszlása a normális eloszlásnál lapultabb.

ábra 4 score hisztogram

A szórás 3.8, azaz egy véletlenszerűen kiválasztott konfiguráció értéke 3.8 ponttal fog eltérni az átlagtól. (Nem beépített R sd, mivel ismerjük a teljes populációt)

Chart, box and whisker chart

Description automatically generatedA konfigurációk negyede több, mint 46 pontos, míg a negyede kevesebb, mint 41 pontos.

Felső kerítés 53.5 (46+1.5\*5). Ez azt jelenti, hogy az ennél nagyobb score értékek kilógóan magasak. A gyakorisági táblából látható 1134 ilyen rekord van (ez az elemek 0.2%-a). Az alsó kerítés 33.5 (41-1.5\*5). Ebben a kategóriában 3696 rekord van (az elemek 0.7%-a).

ábra 5 score dobozábra

## Size (nem numerikus)

Az értékkészlete 3 különböző érték.

A módusz és a medián is „medium”. A medián értelmezhető, mert a három érték egyértelműen sorba rendezhető (ordinális). Ez azt jelenti, hogy konfigurációk leggyakrabban közepes méretű sofőröket tartalmaz, illetve, hogy a sorbarendezett elemek középső értéke is „medium”.

# Intervallumbecslés és Hipotézisvizsgálat

Text

Description automatically generated with medium confidence

Csináltam egy 50 elemű mintát, amivel szimuláljuk egy „háziverseny” összetételét.

A kis méretű sofőrökkel rendelkező konfigurációk 99%-os valószínűséggel legalább 38,8 pontosak lesznek és legfeljebb 44,1. A közepes méretű sofőrökkel rendelkező konfigurációk 99%-os valószínűséggel legalább 41,2, és legfeljebb 45,2 pontosak lesznek. A nagy sofőrök konfiguráció szintén 99%-os valószínűséggel legalább 38,9 és legfeljebb 48,0 lesz. Azt látjuk, hogy a score nem különbözik szignifikánsan, hiszen bármelyik csoport intervalluma fedi a többit.

A vizsgált mintában a score átlaga 42,81.

Feltételezem, hogy a teljes sokasában az átlag legalább 43 (null hipotézis). Alternatív hipotézisem, hogy az átlag kevesebb, mint 43. (right sided test)

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!

[ábra 1 Sofőrök statisztikái (MarioWiki) 4](#_Toc125390539)

[ábra 2 df 5](#_Toc125390540)

[ábra 3 score gyakorisági táblázat 6](#_Toc125390541)

[ábra 4 score hisztogram 6](#_Toc125390542)

[ábra 5 describe score 6](#_Toc125390543)

[ábra 7 score szórás 6](#_Toc125390544)

[ábra 8 score doboz ábra 7](#_Toc125390545)

[ábra 8 score summary 7](#_Toc125390546)