



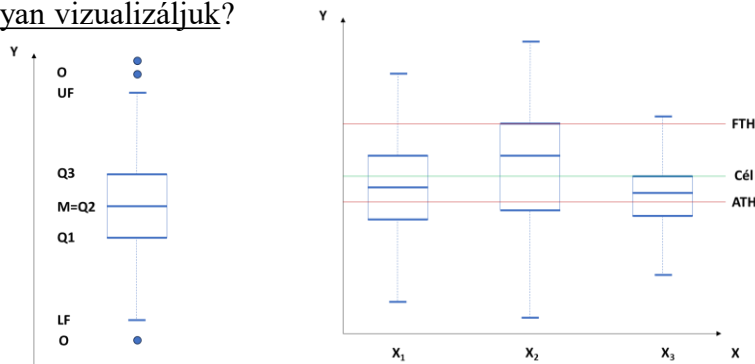
Eszköz neve:

Doboz-ábra (Box Plot)

Mire használjuk?

Adatok eloszlásának néhány statisztikai jellemzőjét összefoglalóan bemutató, egyszerű ábrázolási módszer. Alkalmazható minden olyan esetben, amikor a cél nem a részletek, hanem az eloszlás középértékének és ingadozásának bemutatása, pl. különböző minták összehasonlítása érdekében. Példa: ugyanazt a tevékenységet végző csapatok teljesítményének összehasonlítása, idődiagram (run-chart) adatainak összefoglaló bemutatása.

Hogyan vizualizáljuk?



Mire figyelünk?

- A Q1 az emelkedő sorrendbe tett adatok alsó 50%-ának mediánja, a Q3 pedig a felső 50% mediánja.
- A határok (IF) azokat az adatokat fogják közre, amelyek az eloszlás statisztikailag megbecsülhető halmazát alkotják, adott valószínűségi szinten.
- Az összehasonlításnál a határokon kívüli adatok elhagyhatók, de az elemzésnél figyelembe kell venni ezeket is.

Hogyan használjuk?

1. Rakják emelkedő sorrendbe az adatokat.
2. Határozzák meg az adatsor mediánját (M), vagy más néven 2. kvartilisét (Q2). Páratlan elemszám esetén ez az az érték, amelynél ugyanannyi nagyobb érték van, mint amennyi kisebb. Páros elemszám esetén a két középső érték átlaga.
3. Határozzák meg az adatsor 1. és 3. kvartilisét (Q1, Q3). A Q1 az az érték, amelynél az adatok 25%-a kisebb, 75%-a nagyobb. A Q3 az az érték, amelynél az adatok 75%-a kisebb, 25%-a nagyobb. Páros és páratlan elemszámnál a szabályok megegyeznek a mediánál alkalmazottakkal.
4. Határozzák meg a teljes adatsor középső 50%-ának terjedelmét (range 50 – R50), amely a Q3 és a Q1 különbsége.
5. Határozzák meg az adatsor határait (fences – F) a következő módon:
 - a) a felső határ (upper fence – UF) meghatározása:

$$UF = Q3 + 1,5 R50$$
 - b) az alsó határ (lower fence – IFL) meghatározása:

$$LF = Q1 - 1,5 R50$$
6. Azonosítsák a határokon kívül eső adatokat (outliers – O).
7. Rajzolják meg az ábrát a meghatározott értékekkel.
8. Több eloszlás összehasonlítása esetén készítsenek el egy koordináta rendszert, amelynek X-tengelyén sorolják fel az eloszlások azonosítóit (X₁, X₂ ...), az Y-tengelyen pedig jelölik a mért értékeket.
9. Jelölik vízszintes vonalakkal a mért változóra meghatározott célértéket és a tűréshatárokat.
10. Az összehasonlítás alapja a M, a Q1 és a Q3 célértéktől és tűréshatároktól való eltérésének nagysága és iránya, a doboz szélessége (R50), valamint a határon kívüli adatok mennyisége és értéke.