

## 6. gyakorlat – Követelmények elemzése, felderítő adatelemzés – Megoldások

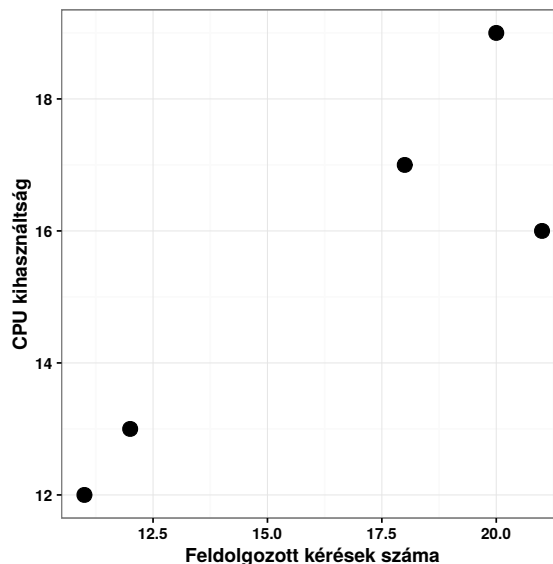
### 1. Szerverteljesítmény felderítő adatelemzése

Egy szerveren az alábbi teljesítményjellemzőket mértük:

Mintavétel időpontja [ms]	500	600	700	800	900
Utolsó 100ms alatt feldolgozott kérések száma [darab]	11	12	21	18	20
Utolsó 100ms átlagos kiszolgálási ideje [ms]	15	20	21	25	27
Utolsó 100ms CPU kihasználtság [%]	12	13	16	17	19
Utolsó 100ms HDD I/O kihasználtság [%]	55	63	87	61	73

- a) Ábrázoljuk a feldolgozott kérések számát és a CPU kihasználtságot pontfelhő (scatterplot) diagramon! Értelmezzük a diagramot!

**Megoldás**



Két klaszter (csoportosulás) látszik, nagyjából pozitív a korreláció (kb. arányosak az adatok), de nem direkt monoton (valami más is befolyásolhatja az adatokat, ezért ingadozik). Értelmezve a látottakat a CPU átlagos kihasználtsága a feldolgozott kérések számával nő. A bal alsó csoport kisebb terhelésű pillanatokot tartalmaz, míg a jobb felső nagyobbakat. Ez a megfigyelés adott esetben jó alapja lehet a terhelés vizsgálatának (pl. az egyes csoportokhoz tartozó pontok időben is közel vannak-e egymáshoz).

- b) Az első mintavétel idején mekkora az átbocsátási ráta értéke? Az 5 mintavétel alapján mekkora az átbocsátási ráta tapasztalati átlaga és mediánja? Mi tartozik a 40%-os kvantilisbe?

**Megoldás**

A mintavételi időkből látszik, hogy két mintavétel között 100 ms telik el. Ebből

$$X_1 = \frac{k_1}{\Delta t} = \frac{11 \text{ kérés}}{100 \text{ ms}} = \frac{11 \text{ kérés}}{100 \text{ ms}} \left[ \frac{1000 \text{ ms}}{1 \text{ s}} \right] = 110 \frac{\text{kérés}}{\text{s}}.$$

A tapasztalati átlag kiszámítása történhet a másik négy átbocsátás kiszámításával és átlagolással, vagy a következő módon (kihasználva, hogy  $\Delta t$  végig 100ms):

$$\bar{k} = \frac{\sum_{i=1}^n k_i}{n} = \frac{11 + 12 + 21 + 18 + 20}{5} = 16,4$$

Ebből az átlagos átbocsátás  $\bar{X} = \frac{\bar{k}}{\Delta t} = \frac{16,4}{0,1} = 164 \frac{\text{kérés}}{\text{s}}$ .

Az elemek sorba állítva 11, 12, 18, 20, 21, ebből rögtön látszik, hogy a medián 18, tehát az átbo-csátás mediánja  $\frac{18}{0,1} = 180 \frac{\text{kérés}}{\text{s}}$ .

A  $p$  kvantilis definíció szerint az a szám, amelynél az elemek  $p$ -ed része kisebb vagy egyenlő. A  $p$  kvantilisba azok az elemek tartoznak, amelyek kisebb vagy egyenlők a  $p$  kvantilishoz. A kvantilis speciálisabb változata a percentilis, amely egész százalékokkal dolgozik, valamint a kvartilis, amely „negyedeli” az adatot. Pl. a 35. percentilis a 35%-os kvantilishoz felel meg (a kvantilis lehetne pl. 35,7% is!), a második kvartilis pedig az 50%-os kvantilishoz.

Itt az elemek legkisebb 40%-a a 11 és a 12, ezért a 40%-os kvantilis értéke a 12 lesz, és a 11, illetve 12 elemek tartoznak bele. A kapcsolódó átbo-csátási ráták  $110 \frac{\text{kérés}}{\text{s}}$  és  $120 \frac{\text{kérés}}{\text{s}}$ .

- c) Vajon mely mért jellemzők között sejthető ok-okozati viszony?

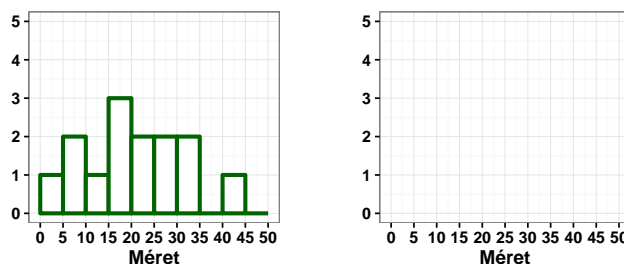
### Megoldás

Ahogy az ábrán is látjuk, az átbo-csátás hatással van az erőforrások kihasználtságra. A szűk keresztmetszetnek számító erőforrás (HDD – ld. korábbi gyak) magas kihasználtsága meg is látszik a megnyúlt válaszdőkön.

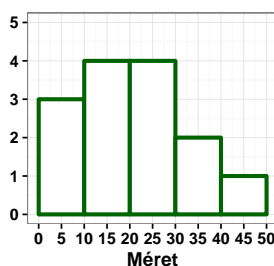
## 2. Képgaléria – adatelemzés

Online képgalériánkban a felhasználók keresés alapján megjeleníthetnek a keresőkifejezésre illeszkedő képeket.

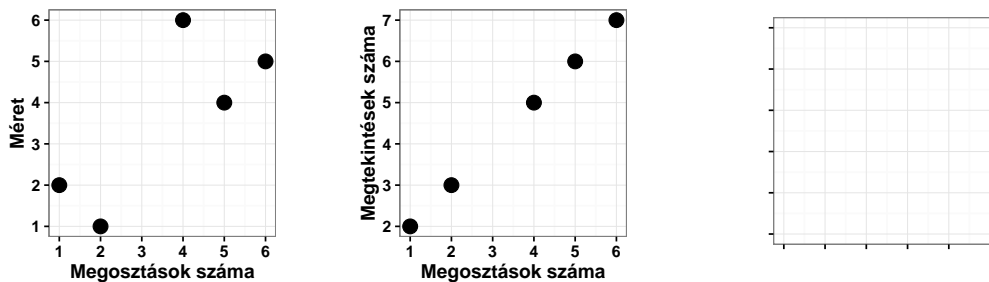
- a) Az alábbi hisztogramon ábrázoltuk az albumok méretének eloszlását. Mivel a tárhely hatékony szervezéséhez elég azt tudnunk, hogy hány 10 alatti, 10 és 20 közötti stb. képet tartalmazó albumunk van, az alábbihoz képest kétszeres oszlopszélességű hisztogramot szeretnénk (szintén a 0 mérettől kezdve felszámítva az oszlopokat). Rajzoljuk meg az ábrát!



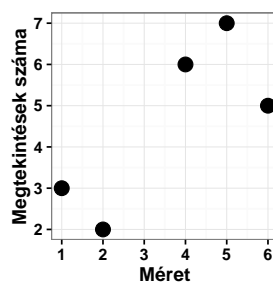
### Megoldás



- b) Pont-pont diagramon (scatterploton) ábrázoltuk 5 kiválasztott album méretét illetve megtekintési számát a megosztási számmal összehasonlításban. Igaz-e, hogy minél nagyobb az album, annál többen tekintik meg? Válaszolja meg a kérdést egy harmadik pont-pont diagramon, amely a megtekintések számát a méret függvényében ábrázolja!



### Megoldás



- c) Az albumok jellemző népszerűségét szeretnénk meghatározni, emiatt a pont-pont diagram alapján kiszámoltuk a megtekintési számok átlagát és mediánját. Általánosságban megtehető-e ez egy pont-pont diagram alapján? Mennyivel változnak ezen középértékek, ha feltöltünk egy új albumot, amelyet 40-en tekintenek meg?

#### Megoldás

Az értékek 2, 3, 5, 6, 7; tehát az átlag  $\frac{23}{5} = 4,6$ , míg a medián 5.

Ha hozzávesszük a 40-et, akkor az átlag  $\frac{63}{6} = 10,5$ , míg a medián  $\frac{5+6}{2} = 5,5$ .

*Tanulság:* az átlag érzékenyebb a kiugró értékekre, a medián kevésbé.

### 3. Vasúti biztosítórendszer követelményelemzése

Vasúti biztosítórendszert tervezünk. A rendszer elsődleges célja a vonatok összeütközésének megakadályozása. A megfelelő rendszer kifejlesztésének kulcsa a jó minőségű követelményspecifikáció, ugyanis a követelmények alapján kell majd teszteseteket és egyéb ellenőrző vizsgálatokat kidolgoznunk.

1. táblázat. A vasúti biztosítórendszer követelményei (részlet)

R1	<b>Biztonság</b>	A felügyelt pályarendszeren tartózkodó vonatok nem ütközhetnek össze.
R2	<b>Működés</b>	A vonatoknak biztosítani kell, hogy elérhessék az úti-céljukat.
R3	<b>Optimalitás</b>	Minimalizálni kell a vonatok menetidejét.
R4	<b>Pályaszakaszok felügyelete</b>	A pályarendszert szakaszokra kell osztani, ezeken egyszerre egy vonat tartózkodhat.
R5	<b>Szakaszokra bontás</b>	A pályarendszert szakaszokra kell bontani.
R6	<b>Foglaltság</b>	Egy szakaszon egyszerre egy vonat tartózkodhat.
R7	<b>Foglaltság érzékelése</b>	Valamilyen módon érzékelni kell, hogy egy szakaszon áll-e vonat, vagy nem.
R8	<b>Hibatűrés</b>	A komponensek meghibásodására fel kell készülni.
R9	<b>Foglaltságjelző szenzorok</b>	A foglaltságot többféle, redundánsan kialakított szenzorral kell érzékelni.
R10	<b>Sínekbe épített szenzor</b>	A sínekbe mindegyik szakaszon szenzorokat kell telepíteni, amik jelzik, hogy a szakaszon áll-e vonat, vagy sem.
R11	<b>Kamerás rendszer</b>	Ahol lehetséges, kamerákat kell telepíteni a szakaszok megfigyelésére.
R12	<b>Helyzetmeghatározás</b>	A vonatoknak folyamatosan jelezni kell a helyzetüket a központi vezérlő felé.
R13	<b>GPS alrendszer</b>	A vonatokat GPS alrendszerrel kell felszerelni.
R14	<b>Vezeték nélküli kapcsolat</b>	Biztosítani kell, hogy a vonatok vezeték nélküli hálózaton jelezhessék a helyzetüket a központi vezérlő felé.
R15	<b>Vonatok vezérlése</b>	Meg kell tudni akadályozni, hogy a foglalt szakaszra másik vonat is ráhajthasson.
R16	<b>Vonat leállítása</b>	A központi rendszer azonnal leállíthatja a vonatot.
R17	<b>Mozdonytípusok támogatása</b>	A rendszernek támogatnia kell minden, a sínpáron közlekedni képes mozdonytípust.
R18	<b>Mozdony nem módosítható</b>	Nem alkalmazható olyan megoldás, amihez a mozdonyok vezérlését meg kellene változtatni.

- a) Gyűjtsük össze azokat a szereplőket, akik egy ilyen rendszer kifejlesztése kapcsán érintettek, vagyis követelményeket támaszthatnak a leendő rendszerrel szemben (ún. *stakeholderek*)!

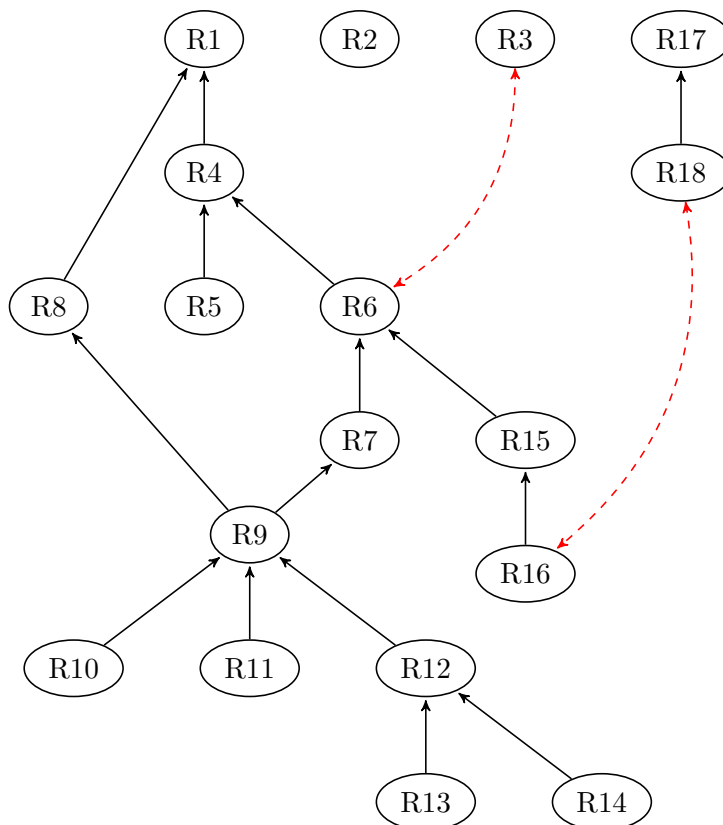
#### Megoldás

- Vasúti társaság
    - Vasúti karbantartók
    - Mozdonyvezetők
    - Forgalmirányítók
  - Vonatgyártók
  - Utasok
  - Hatóságok
    - Törvények és szabályozások
    - Felügyelő testületek
  - Szabványok
  - stb.
- b) A stakeholderek felderítése után összegyűjtöttük az általuk támasztott követelményeket is, ennek egy részletét tartalmazza az 1. táblázat. Rajzoljuk fel a követelmények közötti függőségi viszonyokat egy gráf segítségével! A gráfban  $A$ -ból  $B$ -be mutató irányított éllel jelezzük, ha (1) az  $A$  követelmény a  $B$  követelmény része (*kompozíció*), (2) az  $A$  követelmény finomítja (részletezi) a  $B$  követelményt (*refines* kapcsolat), illetve (3) az  $A$  követelmény származtatható a  $B$  követelményből (*derive* kapcsolat). Ne foglalkozzunk azzal, hogy két követelmény közül ezen viszonyok

melyike áll fent; most csak a kapcsolat megléte a fontos.

### Megoldás

A vázolt kapcsolattípusok alapvetően a követelményanalízis során a tervező top-bottom gondolatmenetét fejezik ki, ezért a kapcsolat megléte, illetve főképp a típusa igen szubjektív lehet. A lényeges információ itt az, hogy adott magas szintű követelményhez mely alacsony szintű követelmények kapcsolódnak valamilyen módon. Egy ilyen hierarchia segítségével a megvalósítás bottom-up módon történhet az alacsony szintű követelményektől indulva. Emiatt nem érdemes most feltüntetni a kapcsolatok típusát.



(A szaggatott piros vonallal jelzett élek későbbi feladatban kerülnek elő.)

- c) A felsoroltak közül melyek funkcionális követelmények, illetve milyen típusúak az extrafunkcionális követelmények (biztonságosság, teljesítmény, megbízhatóság stb.)?

### Megoldás

Funkcionális: R2, R4–R7, R9–R16

Extrafunkcionális: R1 (biztonságosság), R3 („teljesítmény”, esetleg optimalitás), R8 (megbízhatóság), R17 (kompatibilitás), R18 („karbantarthatóság”)

Vegyük észre, hogy egy extrafunkcionális követelmény megvalósításához szükséges konkrét funkciók (származtatott követelmények) már lehetnek funkcionális követelmények; és fordítva, egy funkcionális követelmény teljesülésének lehetnek extrafunkcionális feltételei.

- d) Vizsgáljuk meg, hogy konzisztens-e a bemutatott követelményrendszer! Ha nem az, akkor mutassunk példát ellentmondásra!

### Megoldás

A követelményrendszerben két ellentmondás is található (ezeket az előző ábrán piros szaggatott vonallal jelöltük). Az egyszerűbben feloldható konfliktus az R3 és az R4 (vagy R6) között figyelhető meg, ugyanis azzal, hogy egy szakaszra egyszerre csak egy vonat engedünk be, várhatóan szuboptimális eredményt kapunk a menetidő tekintetében (pl. ha egy szakasz jóval hosszabb, mint egy vonat). Itt kompromisszumot kell kötni a biztonságosság és a hatékonyság között, ami a valóságban is egy gyakori dilemma.

A másik ellentmondás az R16 és R18 között feszül, ugyanis ahhoz, hogy a vonat távolról leállítható legyen a központból, feltételezhetően mindenképpen „bele kell nyúlni” a vezérlőjébe. Ennek a konfliktusnak a feloldása már nem kompromisszum kérdése, itt mindenképpen át kell

alakítani a követelmények egy részét, attól függően, hogy az R16 vagy az R18 megsértése az elfogadhatóbb.

- e) A fentiekből adjunk példát közvetlenül ellenőrizhető követelményre!

### Megoldás

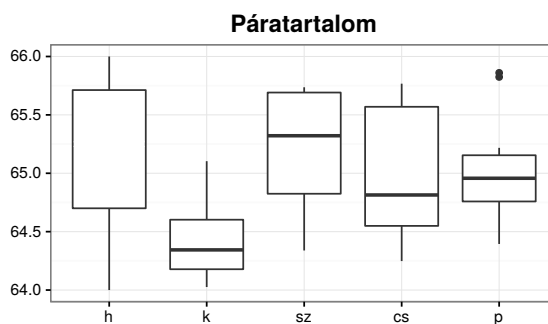
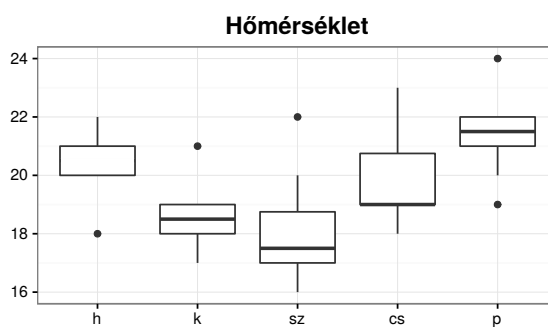
Pl. ilyenek az R6, R7, R9 követelmények.

## 4. Szenzorhálózat (korábbi zh feladat) – adatelemzés

Adott egy mezőgazdasági szenzorhálózat, amellyel a szabadföldes, üvegházi, ill. fóliasátras területeink állapotát követjük nyomon a mért értékek (hőmérséklet, páratartalom, fényerősség, szélesség, detektált kártevők stb.) alapján.

Dátum	Hőm. [°C]	Pára. [%]	Kártevők [db]
2015. 05. 04. 08:00	18	66,00	3
2015. 05. 04. 09:00	20	65,75	6
2015. 05. 04. 10:00	20	65,75	8
2015. 05. 04. 11:00	20	65,50	9
2015. 05. 04. 12:00	20	65,50	5
2015. 05. 04. 13:00	21	65,00	12
2015. 05. 04. 14:00	21	64,70	5
2015. 05. 04. 15:00	21	64,70	6
2015. 05. 04. 16:00	21	64,60	7
2015. 05. 04. 17:00	22	64,00	2

- a) Sajnos a május 4. hétfői középértékek (medián) lemaradtak az ábráról, rajzoljuk őket be a táblázatban található adatok alapján!
- b) Értelmezze a diagramokat: mely változó(k) első kvartilisei mutat(nak) szigorúan monoton változást az idő folyamán?
- c) (Kiegészítő feladat.) Szeretnénk párhuzamos koordináta diagramon összevetni a hétfői hőmérsékleti értékeket a detektált kártevők számával.



### Megoldás



a) Rajzoljuk be a medián értékeket. Mivel páros számú értékünk van, ezért a középső kettő átlaga lesz a medián. Az első két oszlop rendezett, ezért pont a középső két érték átlaga:  $\frac{20+21}{2} = 20,5$ , ill.  $\frac{65,5+65}{2} = 65,25$ .

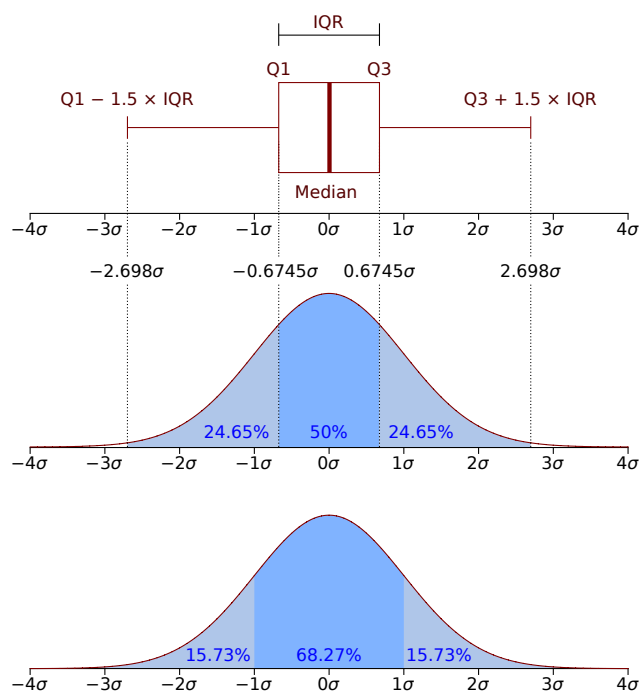
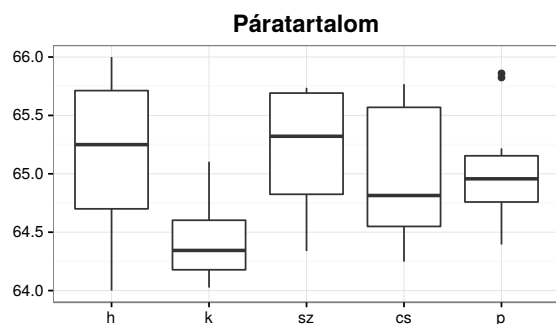
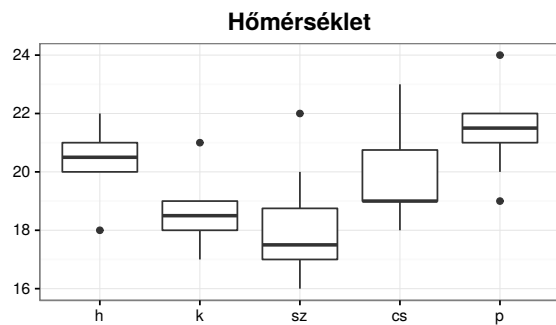
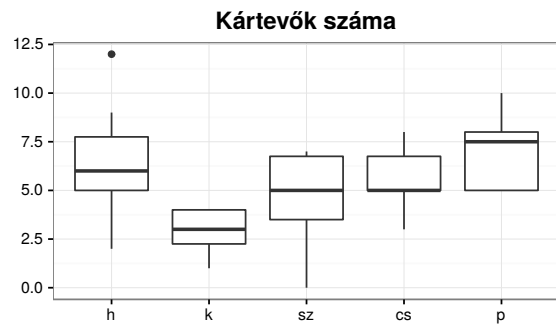
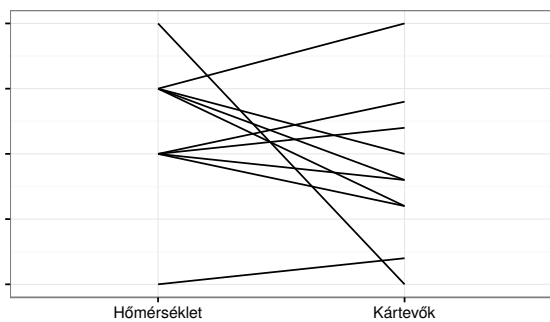
A harmadik oszlop sorbarendezve 2, 3, 5, 5, 6, 6, 7, 8, 9, 12, így a medián  $\frac{6+6}{2} = 6$ .

b) Egyik sem, hiszen a „dobozok” alja nem mutat sehol szigorúan monoton változást.

A boxplot főbb jellemzőit az 1. ábra mutatja be. A  $\pm 1.5 \times IQR$ -en kívül eső értékeket ponttal jelöljük.

Érdekességként megjegyezzük, hogy a 1.5 konstans használata egy statisztikai konvenció, amely analóg a a normális eloszlású adathalmazok  $\pm 3\sigma$  elvével.

c) Az értékeket az alábbi párhuzamos koordináta diagramon ábrázoljuk.



1. ábra. A boxplot főbb jellemzői

## 5. Szenzorhálózat (korábbi zh feladat) – teljesítményelemzés (\*)

(A 4. feladathoz kapcsolódó teljesítményelemzési feladatok.) A különböző típusú szenzorok a helyüktől számított 100 méteres körzetben lévő területekről szolgáltatnak adatokat. A szenzorok mérési eredményeiket időbélyeggel ellátva, rádiós kommunikációs hálózaton továbbítják a központnak. A központi számítógép processzora feldolgozza a kéréseket, majd archiválási cézzattal kiírja őket egy tárolóegységre. A gazdaságunk összesen 4500 szenzort telepített, amelyek percenként egy-egy mérési eredményről adnak jelentést. A rendszer sikerrel kiszolgálja a terhelést. A rádiós kommunikációs hálózat 100 mérési eredményt képes másodpercenként továbbítani. A központi számítógép CPU idejének 75%-a tétlenül múlik. A tárolóegységet 8 ms-ig foglalja le minden egyes kérés kiírása.

a) Másodpercenként hány mérési adat a rendszer jelenlegi átbocsátása?

**Megoldás**

$$X = 4500 \text{ szenzor} \cdot 1 \text{ adat}/60 \text{ s} = 75 \text{ adat}/\text{s}$$

b) Mekkora a hálózat, CPU, ill. tároló átbocsátása, átbocsátóképessége és kihasználtsága?

**Megoldás**

$$X_{\text{hálózat}} = X_{\text{CPU}} = X_{\text{tároló}} = X = 75 \text{ adat}/\text{s}, \text{ mert minden vizitációs szám } 1.$$

$$X_{\text{max}}^{\text{hálózat}} = 100 \text{ adat}/\text{s} \rightarrow U_{\text{hálózat}} = X_{\text{hálózat}}/X_{\text{max}}^{\text{hálózat}} = 75/100 = 75\% = 0,75$$

$$U_{\text{CPU}} = 1 - 0,75 = 0,25 = 25\% \rightarrow X_{\text{max}}^{\text{CPU}} = X_{\text{CPU}}/U_{\text{CPU}} = 75/0,25 \text{ adat}/\text{s} = 300 \text{ adat}/\text{s}$$

$T_{\text{tároló}} = 0,008 \text{ s}$  és nincs átlapolódás:

$$X_{\text{max}}^{\text{tároló}} = 1/T_{\text{tároló}} = 125 \text{ adat}/\text{s} \rightarrow U_{\text{tároló}} = X_{\text{tároló}}/X_{\text{max}}^{\text{tároló}} = 75/125 = 60\% = 0,6$$

c) A mérési pontosság javításához hány szenzort helyezhetünk még üzembe ugyanezen a területen az infrastruktúra fejlesztése nélkül? Feltételezzünk lineáris skálázódást!

**Megoldás**

Mivel minden mérés feldolgozásához igénybe vesszük mindhárom erőforrást:

$$X_{\text{max}} = \min(X_{\text{max}}^{\text{hálózat}}, X_{\text{max}}^{\text{CPU}}, X_{\text{max}}^{\text{tároló}}) = X_{\text{max}}^{\text{hálózat}} = 100 \text{ adat}/\text{s}$$

Tehát 4/3 arányú felskálázás lehetséges, még 1500 szenzor üzembe helyezhető.

d) A rádióhálózat ügyes kódolással biztosítja, hogy egyszerre több szenzor is sugározhasson mérési eredményeket. Átlagosan hány szenzor rádiója sugároz egyszerre (vagyis hány-szoros az átlapolódás) a hálózaton jelenleg, ill. a hálózat maximális terheltsége esetén, ha egy mérési eredmény sugárzása 40 ms-ig tart?

**Megoldás**

Alkalmazzuk Little törvényét a levegőben épp sugárzás alatt álló üzenetekre:

$$T_{\text{hálózat}} = 0,040 \text{ s}$$

$$X_{\text{hálózat}} = 75 \text{ adat}/\text{s} \rightarrow N_{\text{hálózat}} = X_{\text{hálózat}} \cdot T_{\text{hálózat}} = 75 \text{ adat}/\text{s} \cdot 0,04 \text{ s} = 3 \text{ adat egyszerre most}$$

$$X_{\text{max}}^{\text{hálózat}} = 100 \text{ adat}/\text{s} \rightarrow N_{\text{hálózat}} = 100 \text{ adat}/\text{s} \cdot 0,04 \text{ s} = 4 \text{ adat maximálisan}$$